

NUREG/CR-4604
PNL-5849

Statistical Methods for Nuclear Material Management

Date Published: December 1988

Edited by
W. Michael Bowen and Carl A. Bennett

Pacific Northwest Laboratory
Richland, Washington 99352

Prepared for:
**Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN B2420**

Library of Congress Cataloging-in-Publication Data

Statistical methods for nuclear material management.

"Report no.: NUREG/CR-4604 (PNL-5849)."

"Contract no.: FIN-B2420."

Includes index.

1. Nuclear energy--Security measures--Statistical methods.
2. Material accountability--Statistical methods. I. Bowen, W. Michael.
- II. Bennett, Carl A. (Carl Allen), 1921- . III. U. S. Nuclear Regulatory Commission. Office of Nuclear Regulatory Research.

TK9152.S72 1987 519.5'024363--dc19

ISBN 0-87079-588-0

ISBN 0-87079-589-9 (microfiche)

NOTICE

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus product or process disclosed in this book, or represents that its use by such third party would not infringe privately owned rights.

Available as NUREG/CR-4604 from
The Superintendent of Documents
U.S. Government Printing Office
P.O. Box 37082
Washington, DC 20013-7082

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

Price: Paper Copy
Microfiche A01

Printed in the United States of America

1988

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

FOREWORD

This book is intended as a reference manual of statistical methodology for nuclear material management practitioners. It describes statistical methods currently or potentially important in nuclear material management, explains the choice of methods for specific applications, and provides examples of practical applications to nuclear material management problems. Together with the accompanying training manual, which contains fully worked out problems keyed to each chapter, this book can also be used as a textbook for courses in statistical methods for nuclear material management. It should provide increased understanding and guidance to help improve the application of statistical methods to nuclear material management problems.

The U. S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, sponsored the preparation of this book.

Eric S. Beckjord, Director
Office of Nuclear Regulatory Research
U. S. Nuclear Regulatory Commission

PREFACE

Since 1973, the standard manual of statistical analysis methods for nuclear material management has been a book by John L. Jaech, *Statistical Methods in Nuclear Material Control*, which was compiled from material prepared by Jaech for several courses on the application of statistics to problems of nuclear material accounting. The courses were conducted as part of a Safeguards Training Program sponsored by the U. S. Atomic Energy Commission. The book was intended as a personal reference book rather than as a textbook. It did not contain sets of assignment problems or a bibliography but served as a source of practical methods and example problems for self study and supplementary classroom use.

In 1981, the Office of Standards Development of the U. S. Nuclear Regulatory Commission (NRC), subsequently merged with the NRC Office of Nuclear Regulatory Research, contracted with the Pacific Northwest Laboratory, operated by Battelle Memorial Institute for the U. S. Department of Energy, to prepare both a new reference manual and a training manual for guidance in the application of statistical methods in nuclear material management. The reference manual was to be complete enough to serve as a single source of statistical methodology for nuclear material management practitioners. It would describe statistical methods currently perceived as important or potentially important for nuclear material management, provide abundant examples of practical applications to nuclear material management problems, and explain the principles and rationale for the choice of methods in specific applications. The training manual would be a companion volume containing sets of problems for student assignments, together with fully worked out solutions. This book and the companion training manual written by Jaech are the result of this effort.

It should be emphasized that this book, while providing much of the statistical background and methodology needed by nuclear material management practitioners for routine applications, is not always a substitute for professional statistical advice. Statistics is a subtle subject with many potential pitfalls, and there are many situations, sometimes even apparently routine ones, where an experienced statistician should be consulted. It is our hope that this book will help the practitioner master

enough basic statistical methodology to handle many routine applications and to recognize those situations where professional statistical assistance is needed.

W. Michael Bowen
Carl A. Bennett

ACKNOWLEDGMENTS

We wish to acknowledge the contributions of the many individuals who provided suggestions, direction, and other assistance in the preparation of this book. John Jaech of Exxon Nuclear Company, Inc., and Richard Hooper of Battelle Pacific Northwest Laboratories helped to determine the proposed content and critically reviewed some of the first drafts during the initial stages of the project. Dan Lurie of the NRC Office of Administration and Resources Management and John Telford of the NRC Office of Nuclear Regulatory Research not only provided guidance in developing the scope and outline of the book but also continued to provide critical reviews of the draft chapters throughout the project. The NRC project monitors, James Branscome in the earlier stages, and Sandra Frattali and Lee Abramson throughout the remainder of the project, provided constant direction and assistance. Valuable input and constructive suggestions were contributed by the members of the NRC technical review panel: Dan Lurie and John Telford throughout and, at various times, Lee Abramson, Sandra Frattali, Darrel Huff, Jaqueline Kent, Martin Messenger, Mary Miller, and Joseph Rivers. Peer reviewers from several other organizations also made many valuable recommendations that were incorporated into the final manuscript. A special thank you goes to Lee Abramson for his thorough technical and editorial reviews and for his assistance in writing selected portions of the manuscript.

We also thank Mary Lou Lomon, Susan Cammann, Kay Chase, and Sharon Popp for the word processing of the numerous drafts and revisions; Sue Liebetau for her editorial assistance in preparing the manuscript; Hazel Asher of the USDOE Office of Scientific and Technical Information, Oak Ridge, Tennessee, for her assistance as publications manager in the photocomposition of this book; and Kathryn Blackfield and Alice Clayton of Maxima Corporation, Oak Ridge, Tennessee, for the editing associated with the photocomposition.

Finally, we are especially grateful to Richard Brouns, who managed the project through all of its ups and downs, principally authored Chapter 13, read and critiqued the many revisions of every chapter, and provided untiring and invaluable support to the editors at every turn.

W. Michael Bowen
Carl A. Bennett

CONTRIBUTING AUTHORS

The following individuals contributed draft material that was used in preparing the indicated chapters. Except as indicated, all were affiliated with the Pacific Northwest Laboratories at the time of their involvement.

Lee R. Abramson, Nuclear Regulatory Commission, 12
Carl A. Bennett, 13-17
W. Michael Bowen, 2-9, 11
Richard J. Brouns, 13
James W. Johnston, 14
Robert R. Kinnison, 1, 10-12
Judith A. Mahaffey, 1, 10, 11
James A. Merrill, 15
Mark F. Mullen, 17
Gregory F. Piepel, 3, 6-9
Kirkland B. Stewart, 10, 16

INTRODUCTION

Nuclear material accounting is that part of a nuclear material safeguards program that consists of procedures and systems to perform nuclear material measurements, prepare and maintain accounts and records, and perform analyses of the data generated.* The principal method of maintaining and demonstrating control of materials in most industries, including the nuclear industry, is material balance accounting. The essential principle of material balance accounting is maintaining a record of the current (book) inventory of material from input and output data and reconciling it periodically with a physical (measured) inventory. This requires counting and/or measuring all receipts, disbursements and losses of material, maintaining records of these transactions, and periodically counting and/or measuring the quantity of material on hand.

Statistical analysis is an essential element of effective material balance accounting. Although many books are available on general statistical methods, a specialized book on their application to nuclear material accounting can provide methods specific to that application in a single volume and can address some of the unique problems encountered in nuclear material accounting that are not adequately covered in conventional texts. Some of these unique problems are:

1. Multiplicity of error sources in a material balance.
2. Estimation of the variance components associated with each error source.
3. Reconciliation of measurement results from different measurement systems and different laboratories.
4. Independent verification of inventories and material balances.

Statistical methods are needed to help monitor the measurements that are used to generate material accounting data, evaluate accounting anomalies, and determine the effectiveness of the material accounting system. One of the primary uses of statistical methodology is to evaluate uncertainties in the reported amount of nuclear material received, produced,

*For a more complete review of the role of material accounting in nuclear material safeguards see *Report of the Material Control and Accounting Task Force*, 1978, NUREG-0450, Vol. 2, Chapter 1, U. S. Nuclear Regulatory Commission, Washington, D. C.

shipped, lost, discarded, or otherwise added to or removed from the inventory. In addition, statistical methods are used to estimate the effects of these uncertainties on inferences and decisions that are based upon accounting results. The basic issues are the power of the accounting system to detect anomalies and, conversely, the assurance provided by the system that such anomalies are absent and the accounts are correct. For those who possess large amounts of nuclear material, it may be necessary to incorporate tests into a material management system that can detect material losses more promptly than is achieved by a periodic physical inventory. Such tests, which may utilize process yield data and other process information that cross-checks material quantities and locations, also require the use of statistical methodology.

The accounting systems and the periodic material balance tests need to be independently audited on a regular basis to assure the effectiveness of nuclear material management. The audits are commonly referred to as inspections. They generally involve independent tests and measurements to verify the accuracy of the reported inventory data, and they require the use of statistical methods for designing sampling schemes and for evaluating the test and measurement results.

As the title indicates, this book deals with statistical methods for nuclear material management. The discussion until now has been directed toward problems in nuclear material accounting. This emphasis reflects the fact that most of the statistical applications are in the nuclear material accounting area. Nuclear material management, however, also includes other areas requiring statistical treatment, such as auditing and process control, that go beyond nuclear material accounting. Many of the statistical methods presented in this book are applicable to these areas of nuclear material management. Nuclear material management is, however, also closely related to the control of processes for the manufacture of intermediate and final nuclear products. Statistical control charts and other process control techniques are used to control the key chemical and physical product specifications in the fuel production process, and statistical analysis is used to certify finished products that are blends of several individual production batches.

This book presents a unified approach to the solution of these and other problems of concern to the practitioner in the field of nuclear material management. It provides both a guide to the statistical methods needed to support a nuclear material management program and a description of the statistical principles upon which their application is based. It is intended as a reference for the scientists and engineers in research and production facilities who produce and evaluate nuclear material accounting data, and for those in government organizations who develop and administer accountability requirements and audit the accountability programs of facilities that possess nuclear materials.

Most practitioners in this field have had courses in advanced algebra and elementary calculus but only limited formal training in mathematical statistics and data analysis. Therefore, the content and style of the book are directed to those inexperienced in the application of statistical methods. Generally, derivations and statistical theory are not presented. The emphasis is on providing a guide to practice. With this book and a short course in the use of the statistical methods, such individuals will be able to apply many of the methods to the less complex problems that occur in nuclear material management. For the more complex problems, such as modeling and estimating the variances of inventory differences or developing an inventory verification procedure, consultation with a statistician is recommended.

The contents of this book will be valuable to others besides practitioners in nuclear material management. The statistical methods presented can be used in many chemical and engineering applications, including those that deal with nuclear materials. Conversely, the book will help administrators and statisticians become familiar with the material accounting process, particularly through the example problems presented throughout the book and the material accounting overview in Chapter 13. Simplified descriptions of the principal nuclear materials measurements and of a typical accounting system are presented (see the appendixes to Chapter 13).

This book has two parts. Chapters 1 through 12 provide a background in the principles and methods of statistical analysis. They cover such topics as basic probability, estimation, hypothesis testing, analysis of variance, regression analysis, experimental design, statistical sampling, nonparametric methods, simultaneous inference, and detection of outliers. Each chapter contains example problems to help the reader understand and use the statistical methods. Brief introductions to simulation and decision theory are also presented. Chapters 13 through 17 describe specific applications of the statistical methods to nuclear material accounting and demonstrate most of the methods with typical problems.

Although statistical methods currently in regular use are emphasized, some additional statistical tools are included because they have considerable potential value in nuclear material management. Such methods appear particularly in the chapters on experimental design, nonparametric methods, simulation, and decision theory. In a number of cases, methods are not described in detail but are simply summarized, and references are given for further study. In some instances, additional references not specifically referred to in the chapter are also listed. Many of the important methods described in this book are referenced except for some standard statistical methods that can be found in textbooks.

Readers with engineering or scientific training and a good familiarity with nuclear material processes and measurements may gain an improved

understanding of most of the currently used statistical methodology without reading Chapters 1, 2, 7, 9, 11, 12, and 13. Those who need more practice in applying the methods to measurement and accounting data should consult the training manual, which has been published as a companion volume.* A short course based on the training manual with this book as a reference is recommended. For student practice, the training manual presents many statistical problems typical of those that occur in nuclear material accounting.

Most statistical calculations can be performed by computers. Many good statistical software packages are available and more are appearing all the time. No attempt is made in this book to reference or recommend any specific software package. Readers not trained in statistics should apply such computer programs with caution. It is important to understand which algorithms are used in a particular software package to ensure that they are applicable to the problem. For the most part, hand calculations are shown for the problems illustrated in this book to illustrate clearly how they are solved. The number of digits retained in the results is often more than is significant, but this retention is done to minimize error in subsequent calculations due to rounding or truncating intermediate results. If hand-held calculators are used, the number of figures that the calculator can retain in memory may sometimes be too limited for accurate calculations, especially where small differences between large numbers are involved. The user should be always aware of this possibility. If the reader cannot get exactly the same answer to a problem as presented in this book, the difference may be due to the rounding or truncating effect.

When the example problems given in this book are based upon data drawn from other publications, reference is made to the source. If no such reference is made, it should be assumed that the problem and its data were synthesized, generally by a simulation, or were taken from the author's personal experience. It is not intended that any set of data represents any specific facility or occurrence.

The authors have generally adhered to the recommended notation and terminology of the draft 1982 ANSI standard.† When additional notation or terminology is necessary, the terms are defined when introduced. One of the departures from common terminology occurs in Chapter 8, Statistical Sampling of Finite Populations. In that chapter, the terminology common in the survey sampling literature rather than that in the traditional statistical literature is used. Another terminology issue occurs with respect to the usage of the terms *short-term systematic error variance* and *long-term*

*Jaech, J. L., 1988, *Training Manual for Statistical Methods for Nuclear Material Management*, NUREG/CR 4605 (PNL-5855), U. S. Nuclear Regulatory Commission, Washington, D. C.

†American National Standards Institute, Inc., *Statistical Terminology and Notation for Nuclear Materials Management*, ANSI N15.5-1982 (draft), New York.

systematic error variance by Jaech* and others in the nuclear material safeguards literature. These terms have been applied for many years to describe the variance components associated with effects which are common to some subset or all of the results in an inventory difference or one of its components. In this book, the traditional statistical terminology is adhered to in describing such error sources. The need for identifying the various error effects in a model of the material accounting process first occurs in Chapter 14, where in Sections 14.2.3, 14.3.1, and 14.3.3 the concepts of the effects of calibrations, bias corrections, and other sources of error are defined.

*Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, National Technical Information Service, U. S. Department of Commerce, Springfield, Virginia.



CONTENTS

	Page
1. DATA DESCRIPTION AND DISPLAY	1
1.0 Introduction	1
1.1 Scales of Measurement	2
1.1.1 Nominal Scale	3
1.1.2 Ordinal Scale	3
1.1.3 Interval Scale	4
1.1.4 Ratio Scale	4
1.2 Descriptive Statistics: Measures of Central Tendency	4
1.2.1 Mean (Arithmetic Mean)	5
1.2.2 Weighted Mean	6
1.2.3 Median	7
1.2.4 Mode	8
1.2.5 Geometric Mean	8
1.3 Descriptive Statistics: Measures of Variability	8
1.3.1 Range	9
1.3.2 Variance and Standard Deviation	9
1.3.3 Geometric Standard Deviation	10
1.3.4 Percentiles	11
1.4 Data Display: Single Variable	11
1.4.1 Pie Charts	11
1.4.2 Bar Charts	12
1.4.3 Histograms	12
1.4.4 Stem-and-Leaf Displays	15
1.4.5 Boxplots	17
1.5 Data Display: Multiple Variables	19
1.5.1 Scatterplots	20
1.5.2 Time Series Plots	21
1.6 Data Smoothing	22
1.6.1 Running Medians	22
1.6.2 Running Averages	26
1.6.3 Smoothing the Endpoints	26
1.6.4 Examining the Residuals	27

2. PROBABILITY	29
2.0 Introduction	29
2.1 Basic Probability Concepts	29
2.2 Random Variables and Probability Distributions	36
2.3 Expected Values	41
2.4 Certain Discrete Distributions	44
2.4.1 Hypergeometric Distribution	44
2.4.2 Binomial Distribution	47
2.4.3 Poisson Distribution	50
2.5 Certain Continuous Distributions	53
2.5.1 Uniform Distribution	53
2.5.2 Normal Distribution	56
2.5.2.1 Central Limit Theorem	59
2.5.2.2 Normal Approximation to the Binomial	62
2.5.3 Lognormal Distribution	64
2.5.4 Other Continuous Distributions	67
2.6 Jointly Distributed Random Variables	69
2.6.1 Discrete Bivariate Distributions	70
2.6.2 Continuous Bivariate Distributions	73
2.6.3 More Than Two Variables	76
2.6.4 Two Special Multivariate Distributions	79
2.6.4.1 Multinomial Distribution	79
2.6.4.2 Bivariate Normal Distribution	82
2.7 Functions of Random Variables	88
2.7.1 Linear Combinations of Random Variables	88
2.7.2 General Functions of Random Variables	90
3. ESTIMATION	95
3.0 Introduction	95
3.1 Empirical Distributions	96
3.1.1 Empirical Density and Cumulative Distribution Functions	96
3.1.2 Sample Moments	97
3.2 Properties of Point Estimators	98
3.3 Methods for Deriving Estimators	99
3.3.1 Method of Moments	99
3.3.2 Least Squares	100
3.3.3 Maximum Likelihood	100
3.3.4 Bayesian Estimation	102
3.4 Point Estimators	104
3.4.1 Estimators for the Mean and Variance of a Normal Distribution	104
3.4.2 Estimators for the Binomial Distribution Parameter	105

3.4.3	Estimators for Other Distributions	106
3.5	Sampling Distributions and Interval Estimators	106
3.5.1	Variances and Ratios of Variances for the Normal Distribution	108
3.5.2	Means and Differences of Means for the Normal Distribution	112
3.5.3	Interval Estimators for Binomial Distribution Parameters	117
3.5.4	Interval Estimators for the Hypergeometric Distribution Parameter D	122
3.6	Estimators for Functions of Random Variables	124
3.6.1	Point Estimators	124
3.6.2	Interval Estimators for Means	127
3.7	Tolerance Limits for Normal Distributions	129
3.7.1	Two-Sided Tolerance Limits	129
3.7.2	One-Sided Tolerance Limits	131
4.	HYPOTHESIS TESTING	135
4.0	Introduction	135
4.1	Basic Concepts	135
4.1.1	Statistical Hypotheses	135
4.1.2	Type I and Type II Errors	137
4.2	Single Sample from a Normal Distribution	139
4.2.1	Tests About the Mean, Variance Known	139
4.2.1.1	Two-Sided Tests	139
4.2.1.2	One-Sided Tests	142
4.2.1.3	Probability of Type II Error and Power	143
4.2.1.4	Sample Size	147
4.2.1.5	Confidence Intervals and Hypothesis Testing	151
4.2.2	Tests About the Mean, Variance Unknown	152
4.2.3	Tests About the Variance	157
4.3	Independent Samples from Two Normal Distributions	161
4.3.1	Comparison of Two Means, Variances Known	161
4.3.2	Comparison of Two Means, Variances Unknown	165
4.3.2.1	Comparison of Two Means, Assuming Equal Variances	165
4.3.2.2	Comparison of Two Means, Assuming Unequal Variances	165
4.3.3	Comparison of Two Variances	171
4.4	Comparison of Two Means Using Paired Observations	176
4.5	Tests on Proportions and Count Rates	182
4.5.1	Exact Tests on the Binomial Parameter p	183
4.5.2	Approximate Tests on the Binomial Parameter p	188

4.5.3	Comparing Two Proportions from Independent Samples	191
4.5.4	Exact Tests on the Poisson Parameter	195
4.5.5	Approximate Tests on the Poisson Parameter	199
4.6	Sequential Tests	202
4.7	Control Charts	210
4.7.1	Theoretical Control Charts	211
4.7.1.1	Theoretical Control Charts for Means, Proportions, and Rates	211
4.7.1.2	Theoretical Control Charts for the Variance	215
4.7.1.3	Theoretical Control Charts for the Range	216
4.7.2	Empirical Control Charts	217
4.7.2.1	Empirical Control Charts for Means, Proportions, and Rates	218
4.7.2.2	Empirical Control Charts for the Range	223
4.7.3	Cumulative Sum Charts	224
5.	ANALYSIS OF VARIANCE	233
5.0	Introduction	233
5.1	One-Way Fixed Effects Analysis of Variance	234
5.1.1	An Overall Test for Equal Means	234
5.1.2	Unequal Sample Sizes	242
5.1.3	Power and Required Sample Size	244
5.1.4	Estimation and Comparison of Class Means	248
5.1.4.1	Confidence Intervals for the Class Means	249
5.1.4.2	Planned Comparisons	252
5.1.4.3	Duncan's Multiple Range Test	256
5.1.4.4	Comparisons Suggested by the Data	261
5.2	One-Way Random Effects Analysis of Variance	264
5.2.1	Variance Components	265
5.2.2	Power and Required Sample Size for Testing $H_0: \sigma_\alpha^2 = 0$	270
5.2.3	Estimation of the Overall Mean μ	274
5.2.4	Sampling to Estimate μ	279
5.3	Analysis of Variance with Nested (Hierarchical) Models	282
5.3.1	The Nested Random Effects Model	283
5.3.1.1	Variance Components: Estimation and Hypothesis Tests	283
5.3.1.2	Estimation of the Overall Mean μ	291
5.3.1.3	Equal Sample Sizes at Each Stage	298
5.3.2	The Nested Mixed Effects Model	299

5.3.2.1	An Overall Analysis of Variance	300
5.3.2.2	Estimation and Comparison of the Means $\mu_i = \mu + \alpha_i$	301
5.3.2.3	Equal Sample Sizes at Each Stage	308
5.4	Analysis of Variance with Two-Way Crossed Classifications	310
5.4.1	Computational Formulas	310
5.4.2	The Fixed Effects Model	313
5.4.3	The Random Effects Model	320
5.4.4	The Mixed Effects Model	324
6.	REGRESSION ANALYSIS	331
6.0	Introduction	331
6.1	An Overview of Regression and Correlation	331
6.1.1	Correlation	332
6.1.2	The Regression Process	333
6.2	Simple Linear Regression Basics	336
6.2.1	An Example	336
6.2.2	Model and Error Structure	338
6.2.3	Ordinary Least Squares	338
6.2.4	Other Regression Models	341
6.2.4.1	Simple and Multiple Regression Equations	342
6.2.4.2	Transformations of the Dependent Variable	342
6.2.4.3	Errors in Both Variables	343
6.3	General Multiple Linear Regression	344
6.3.1	The General Linear Regression Model	344
6.3.2	Parameter Estimation	346
6.4	Evaluating a Fitted Regression Equation	347
6.4.1	Regression Analysis of Variance	348
6.4.2	F-Test for Significance of Regression	350
6.4.3	Sampling Distribution of $\hat{\beta}$	352
6.4.4	Testing for Lack of Fit	356
6.4.5	Examining Residuals	362
6.4.5.1	Plotting Residuals	363
6.4.5.2	Standardized Residuals	370
6.4.5.3	Serial Correlation	373
6.4.6	Detection of Outliers and High Leverage Observations	378
6.4.6.1	Detecting High Leverage Points	379
6.4.6.2	Measures of Influence	380
6.4.7	Model Validation	382
6.5	Using a Fitted Equation	383

7. EXPERIMENTAL DESIGN	387
7.0 Introduction	387
7.1 Experimental Design Basics	388
7.2 Designs for Single Factor Experiments	391
7.2.1 Completely Randomized Designs	391
7.2.2 Randomized Complete Block Designs	395
7.2.3 Latin Square Designs	398
7.2.4 Other Designs	403
7.3 Factorial Designs for Multifactor Experiments	403
7.3.1 Effects and Interactions	403
7.3.2 2^k and 3^k Designs	408
7.3.2.1 2^k Designs	409
7.3.2.2 3^k Designs	416
7.3.3 Mixed Level Factorial Designs	419
7.4 Fractional Factorial Designs	419
7.4.1 Fractions of 2^k Designs	421
7.4.2 Fractions of 3^k Designs	427
7.4.3 Fractions of General Factorial Designs	427
7.4.4 Blocking with Fractional Replicates	427
7.5 Multifactor Experiments with Randomization Restrictions	430
7.5.1 Randomized Complete Block and Latin Square Designs for Multifactor Experiments	430
7.5.2 Split-Plot Designs	430
7.6 Nested (Hierarchical) Designs	433
 8. STATISTICAL SAMPLING OF FINITE POPULATIONS	 439
8.0 Introduction	439
8.1 Basic Sampling Techniques	441
8.2 Developing a Sampling Plan	443
8.3 Sampling Plans for Estimating Means and Totals	447
8.3.1 Simple Random Sampling Plans	449
8.3.1.1 Limitations of Finite Population Sampling Methods	452
8.3.2 Cluster Sampling Plans	454
8.3.2.1 Simple One-Stage Cluster Sampling	455
8.3.2.2 Simple Two-Stage Cluster Sampling	458
8.3.3 Probability Proportional to Size (PPS) Sampling Plans	460
8.3.3.1 PPS Sampling with Replacement	461
8.3.3.2 PPS Sampling Without Replacement	467
8.3.4 Stratified Sampling Plans	473
8.3.4.1 How to Stratify a Population	474

8.3.4.2 Stratified Random Sampling	476
8.3.4.3 Other Stratified Sampling Plans	480
8.3.5 Confidence Intervals for Population Means and Totals	481
9. NONPARAMETRIC METHODS	483
9.0 Introduction	483
9.1 Single Sample or Matched Pairs	484
9.1.1 A Test of Randomness: The Runs Test	484
9.1.2 Single Sample Tests of Location	489
9.1.2.1 The Sign Test	489
9.1.2.2 The Wilcoxon Signed Ranks Test	491
9.1.3 Matched Pairs Tests for Comparing Locations	493
9.2 Two Independent Samples	495
9.2.1 The Mann-Whitney Test for Comparing Locations	495
9.2.2 The Squared Ranks Test for Comparing Variances	498
9.3 Several Independent Samples	501
9.3.1 Tests for Comparing Locations	501
9.3.1.1 The Median Test	501
9.3.1.2 The Kruskal-Wallis Test	503
9.3.2 The k-Sample Squared Ranks Test for Comparing Variances	507
9.4 Several Related Samples	510
9.4.1 Comparison of Locations for Randomized Complete Block Designs	510
9.4.1.1 The Quade Test	510
9.4.1.2 The Friedman Test	513
9.4.1.3 The Rank ANOVA Test	516
9.4.2 Comparison of Locations for Other Designs	517
9.5 Measures of Rank Correlation	518
9.5.1 Spearman's Rho	518
9.5.2 Kendall's Tau	521
9.6 Goodness-of-Fit Tests	523
9.6.1 Chi-Square Test	523
9.6.2 Kolmogorov Test	529
9.6.3 Tests for Normality	531
9.6.3.1 The Lilliefors Test for Normality	531
9.6.3.2 The Shapiro-Wilk Test for Normality	532
10. SIMULTANEOUS INFERENCE AND DETECTION OF OUTLIERS	537
10.0 Introduction	537

10.1	The Simultaneous Inference Problem	538
10.1.1	Error Rates	538
10.1.2	Allocation of Error	540
10.2	Simultaneous Inference Methods	541
10.2.1	Repeated Normal Statistics	542
10.2.2	Bonferroni t Statistics	543
10.2.3	Studentized Maximum Modulus	544
10.2.4	Chi-Square Sums	544
10.2.5	Fisher's Method	546
10.3	Multivariate Statistical Methods	547
10.3.1	Hotelling's T^2 -Test	547
10.3.2	Extensions of Hotelling's T^2 -Test	552
10.4	Test for Outliers	552
10.4.1	Methods of Dixon	554
10.4.2	Grubb's Treatment of Outliers	556
10.4.3	Studentized Residual Approach	560
11.	SIMULATION	569
11.0	Introduction	569
11.1	The Role of Simulation	570
11.2	Considerations in Simulation Modeling	571
11.2.1	Types of Models	572
11.2.2	Development of a Model and Study Plan	575
11.2.3	Operational Activities	576
11.3	Generation of Random Deviates	576
11.3.1	Random Numbers	577
11.3.2	The Uniform Distribution	579
11.3.3	The Normal Distribution	582
11.3.4	The Chi-Square Distribution	587
11.3.5	The Central and Noncentral Student's t Distributions	589
11.3.6	The F Distribution	592
11.3.7	The Multivariate Normal Distribution	593
11.3.8	The Binomial Distribution	594
11.3.9	Functions of Random Variables	595
12.	ELEMENTS OF DECISION THEORY	601
12.0	Introduction	601
12.1	Decision Analysis	602
12.2	Definitions	602
12.3	An Illustrative Example	603
12.3.1	Statement of the Problem	603
12.3.2	Choosing a Strategy	605
12.3.3	Weighting of Losses and Bayes Strategies	608
12.3.4	Admissible Strategies	610

12.3.5 Mixed Strategies	610
12.3.6 Choosing a Bayes Strategy	612
12.3.7 The Regret Approach	615
12.4 Discussion	616
13. NUCLEAR MATERIAL ACCOUNTING	619
13.0 Introduction	619
13.1 The Material Accounting Process	619
13.1.1 Basic Material Accounting Practices	619
13.1.2 Nuclear Material Accounting	621
13.2 Implementation of Nuclear Material Accounting	622
13.2.1 Facility Design and Process Flow	623
13.2.2 The Accounting System	627
13.2.3 Measurement Systems	629
13.3 The Role of Statistics	630
13.3.1 The Effect of Measurement Uncertainty	631
13.3.2 Estimating and Controlling Measurement Uncertainty	632
13.3.2.1 Measurement Control Programs	632
13.3.2.2 Estimating Measurement Variability	634
13.3.3 Inferences from Material Accounting	636
Appendix 13A The Accounting System	639
13A.1 Introduction	639
13A.2 Receipts	640
13A.3 Internal Transfers	641
13A.4 Shipments	641
13A.5 Waste Discards	641
13A.6 Physical Inventories	643
13A.7 Inventory Adjustments	644
13A.7.1 Inventory Differences and SRDs	644
13A.7.2 Measurement and Bookkeeping Mistakes and Bias Corrections	645
13A.7.3 Radioactive Decay and Fission or Transmutation	646
13A.7.4 Change of Enrichment	646
Appendix 13B Measurement Systems	647
13B.1 Introduction	647
13B.2 Mass Measurements	647
13B.3 Volume Measurements	650
13B.4 Flow Measurements	652
13B.5 Chemical Analysis	653
13B.5.1 Determining Uranium Concentrations	653
13B.5.2 Determining Plutonium Concentrations	656
13B.5.3 The Isotopic-Dilution Method of Analysis	657

13B.6 Sampling	659
13B.7 Nondestructive Assay	661
13B.8 Isotopic Analysis	663
14. MODELING MEASUREMENT VARIABILITY	665
14.0 Introduction	665
14.1 Modeling Measurement Systems	665
14.1.1 The Basic Approach	665
14.1.2 The Accounting Process	669
14.1.3 The Error Structure	670
14.2 Measurement Errors	672
14.2.1 Random Errors and Biases	672
14.2.2 Errors in Item Amounts	675
14.2.3 Errors in Stratum Totals	678
14.3 Expected Values and Variances of Accounting Results	683
14.3.1 Assumed Properties of Error Effects	683
14.3.2 Expected Value and Variance of Stratum Totals	686
14.3.3 Calculating the Variance of Stratum Totals	689
14.3.4 Variance of the Variance Estimator	698
14.3.5 Combining Variances for Strata and Components	700
Appendix 14A An Example of Modeling Measurement Variability	713
14A.1 Introduction	713
14A.2 Measurement System Description	713
14A.3 Modeling the Weighing Operation	716
14A.3.1 Constructing the Error Model	716
14A.3.2 Combining Errors and Taking Expectations	719
14A.3.3 Defining the Variance Components	725
14A.3.4 Calculating the Total Variability	728
15. ESTIMATING AND CONTROLLING MEASUREMENT VARIABILITY	731
15.0 Introduction	731
15.1 Calibration of Measurement Systems	732
15.1.1 Single-Point Calibration of NDA Instruments	732
15.1.2 General Linear Calibration	737
15.1.2.1 Constant Variance in the Measured Responses	737
15.1.2.2 Nonhomogeneity of Variance in the Responses	747
15.1.2.3 Recalibration Procedures	752

15.1.3	Nonlinear Calibration	760
15.1.3.1	Simple Nonlinear Calibration	760
15.1.3.2	Multiple Data Sets	767
15.1.4	The Cumulative Error Model	774
15.1.5	Inverse Calibration	779
15.2	Measurement Control Procedures	780
15.2.1	Monitoring Measurement Processes	
	Against Standards	781
15.2.1.1	An Example of Routine Monitoring	781
15.2.1.2	Continuous Monitoring Against a Standard	790
15.2.1.3	Procedures Involving Several Standards	793
15.2.2	Interlaboratory Control Programs	799
15.2.2.1	Shipper-Receiver Bias	800
15.2.2.2	Interlaboratory Exchanges	807
15.3	Estimates from Repeated Measurements	812
15.3.1	Data from a Single Source	812
15.3.2	Data from Two Sources	815
15.3.3	Data from Three or More Sources	822
15.3.3.1	Grubbs Approach	823
15.3.3.2	Analysis of Variance Approach	826
16.	EVALUATION OF ACCOUNTING RESULTS	831
16.0	Introduction	831
16.1	An Overview of the Evaluation Problem	831
16.1.1	Scope of the Problem	831
16.1.2	Defining the Components of an Inventory Difference	833
16.1.3	Loss Models	849
16.2	Testing Inventory Differences and Cumulative Inventory Differences	850
16.2.1	Hypothesis Testing	851
16.2.2	Decision Theoretic Approaches	859
16.3	Specialized Estimation and Testing Procedures	861
16.3.1	Specialized Loss Estimators	861
16.3.2	Sequential Tests	866
17.	VERIFICATION OF ACCOUNTING DATA	871
17.0	Introduction	871
17.1	The Verification Process	871
17.1.1	The Internal Control System	872
17.1.2	Fundamental Concepts	874
17.1.3	Selecting a Random Sample	877
17.2	Inferences from Observation of Attributes	879

17.2.1 Attribute Sampling	880
17.2.2 Applications of the Hypergeometric Distribution	881
17.2.3 Applications of the Binomial Distribution	888
17.3 Inferences from Measurement Differences	894
17.3.1 Inferences About Measurement Errors	894
17.3.1.1 Use of Inspection Measurements to Verify Measurement Capability and Performance	895
17.3.1.2 Evaluating Inspection Data to Assess Measurement Performance	896
17.3.1.3 Determining the Sample Size	901
17.3.2 Inferences About Material Quantities	903
17.3.2.1 Definition of the Difference Statistic	903
17.3.2.2 The Variance of the Difference Statistic	906
17.3.2.3 The Variance of the Corrected Inventory Difference	909
17.3.2.4 Variables Sampling Plans	919
APPENDIX A STATISTICAL TABLES	927
INDEX	987

CHAPTER 1

Data Description and Display

1.0 INTRODUCTION

Statistics is the science and the art of collecting, organizing, summarizing, analyzing, interpreting, and presenting data to understand the information that the data contain and to reach valid conclusions from the data. Thus, data may be used for **descriptive** or for **inferential** purposes. For example, data on the output of a fuel production facility may be used to describe the level of production or to infer whether the level of production has changed. Often data are used for both purposes.

When data are being collected on characteristics of objects or events, such as the uranium content of nuclear fuel pellets, it is usually impractical to measure the entire **population** or **universe**, the totality of all actual or conceivable objects or events. Instead, a part of the population, called a **sample**, is examined. If items in a sample are selected so that all possible samples of size n have an equal chance of being chosen, the sample is a **random sample** of size n . To make valid inferences about a population from a sample, it must be a random sample. Methods for developing sampling plans for specific purposes and for making inferences from the resulting data that are collected are the topics of Chapters 2 through 12.

Most features of a population are described by numerical constants called **parameters**. A measurable characteristic of an object or event is called a **variable**. When a population is sampled, **observed values** (or measurements) of a variable are examined. Parameters are estimated from observed values of variables. If all possible observed values of a variable in a population were known, the parameter for the feature of interest could be calculated exactly. This is usually not the case, however. Thus, statistical theory and methods are applied to describe and make inferences about parameters based on the observed values of a variable in a random sample. Methods that summarize the information in a random sample are presented in this chapter. In later chapters, statistical methods are applied to the observed values of variables to estimate and to make inferences about parameters. Chapter 2 includes formal definitions, discussions, and examples of populations, parameters, and variables.

If a variable theoretically can assume any value within some interval, it is **continuous**. Fuel pellet weight is a continuous variable. A variable is **discrete** if it can assume only a countable or finite number of values in the

interval. Counts of alpha emissions from fuel pellets are discrete variables. Even though a variable is continuous in theory, it may be measured in such a way that the observed values are discrete. Chapter 2 includes more information on different types of variables.

A **statistic** is a number calculated from observed values of one or more variables. For example, the average weight observed in a sample of fuel pellets is a statistic. The plural of statistic has a different meaning than the same term used above to denote a scientific discipline; that is, statistics are numbers calculated from observed values of variables. Statistics (the plural of statistic) that describe a sample without making inferences about the population are one of the topics of this chapter.

The purpose of descriptive methods is to reveal characteristics of a sample which may be concealed by the collection of numbers. Descriptive methods are both mathematical and graphical. Descriptive methods summarize and display the data so that the structure can guide the choice of inferential methods. Descriptive methods suggest features of the sample and, thus, of the population, but do not confirm their existence as do inferential methods. The use of one or more descriptive techniques to explore and understand the information in data is called **exploratory data analysis**.

Data exploration is a first step in statistical analysis, but exploratory techniques can be useful at all phases of analysis, as new information is revealed and explored. In addition, exploratory data analysis may suggest valid assumptions and lead to the use of appropriate inferential statistical methods. Confirmatory data analyses use data to make inferences about populations, and graphical techniques may be used to infer whether assumptions are reasonable. Thus, methods presented in this chapter for displaying and summarizing data may be related to confirmatory techniques presented in later chapters. In this chapter, however, graphical techniques for information display relate to summarizing and describing data.

Many excellent computer software packages are available that have been specifically developed for data display and/or exploratory data analysis. Utilizing such a software package, especially with large data sets, can simplify implementation of the techniques presented in this chapter and can greatly reduce the required time and effort. Many such software packages also provide techniques that are not presented in this book.

1.1 SCALES OF MEASUREMENT

A **measurement** (or observed value) results from assigning numbers or labels to objects or events. The numbers or labels contain information about a characteristic of the objects or events. The way in which such numbers or labels are assigned imposes restrictions on the use of the measurements in data analysis, and these restrictions cannot be violated if

valid inferences are to be made. Four measurement scales are used to describe measurements, and these relate to the permissible arithmetic operations with the measurements. The scales are nominal, ordinal, interval, and ratio.

1.1.1 Nominal Scale

When numbers or symbols are used to identify groups (or categories) to which objects belong, the scale of measurement is said to be **nominal**. Numbers are used only as a label. In addition to group identification, when nominal data are summarized, the summary includes an identification and a count of the number of items within each group. These counts are called **frequencies**. Frequencies can only assume integer values; there cannot be 2.5 nuclear facilities in a region, but there may be 2 or 3.

Suppose that the storage area of a nuclear fuel facility contains red and yellow storage drums. The number of each color of drum is the corresponding frequency, and a table showing the number (or frequency) of each color gives an inventory of the storage area. Drum color can represent drum content, such as red for scrap and yellow for waste. On the other hand, drums can be labeled by an **identifier** other than color to represent drum content. The identifier 1 could represent scrap; 2, feed material; 3, pellets; and 4, wastes. Arithmetic operations may be performed on the frequencies but not on the numerical identifiers. Frequencies may be statistically compared to some prescribed or expected value. Both frequencies and identifiers are discrete numbers; however, identifiers are qualitative, whereas frequencies are quantitative.

1.1.2 Ordinal Scale

If, in addition to the grouping that characterizes nominal measurements, a relative ordering or ranking of groups can be identified (but not necessarily quantified), the measurements are **ordinal** in scale. Ordinal relationships are typically assigned consecutive integral numbers as identifiers. These identifiers are called **ranks**. Rankings are discrete and quantitative. An order preserving transformation of ranks does not change the information contained in the ranks. That is, it does not matter whether the rank of 1 is assigned to last place, 2 to second-to-last and on up; or whether the rank of 1 is assigned to first place and groups are ranked in descending order.

Suppose that the storage area of a nuclear fuel facility contains large (55-gal), medium (30-gal), and small (5-gal) drums. The capacities of these drums can be identified by (L, M, S), (55, 30, 5), (A, B, C), or by the ranks (1, 2, 3) or (3, 2, 1). The scale applies to the drum capacities, not to contents. (If the contents are unknown then the ordinal scale for the contents is not necessarily the same as for the capacities.)

1.1.3 Interval Scale

When measurements have characteristics of an ordinal scale and, in addition, the interval size (distance) between objects can be quantified, the measurements are said to be on an **interval scale**. An interval scale is characterized by a unit of measurement that assigns a distance between all pairs of objects. A transformation or arithmetic operation on interval values preserves the ordering and relative distances between objects. The differences may be discrete, e.g., if the measurements are counts. The interval scale may be either continuous or discrete. The interval scale is sometimes referred to as a metric scale.

Temperature is an example of an interval scale measurement. Celsius and Fahrenheit scales demonstrate the arbitrary nature of both the zero point and the distance defined to be a unit. Thus, for a given time period, the ratios of daily maximum to minimum temperature differ for Fahrenheit and Celsius temperature scales. As another example, the liquid level (in inches or centimeters) in a tank is an interval scale measurement. It is an interval scale because there is an arbitrary heel below the zero level that may vary from tank to tank.

1.1.4 Ratio Scale

When measurements have the characteristics of an interval scale and also a physically definable zero point, the measurements are said to be on a **ratio scale**, where **zero** is the number that defines the absence of a quantity. For example, measurements of weight, length, and volume are on a ratio scale. For this scale, the ratio of any two measured values of the same type is independent of the unit of measurement. For example, the ratio of height to diameter of a cylindrical tank is the same whether English or metric units are used. (The ratio of length to volume *will* depend on the units). Like the interval scale, the ratio scale may be either continuous or discrete.

1.2 DESCRIPTIVE STATISTICS: MEASURES OF CENTRAL TENDENCY

Sections 1.2 and 1.3 discuss statistics that describe objects or events. For any scale of measurement, certain descriptive statistics may be obtained by grouping the observed values of a variable into categories. For all scales except the nominal scale, there are statistics that may be obtained by ordering the observed values in increasing (or decreasing) order of magnitude. For interval and ratio scale measurements, there are many statistics that are obtained through arithmetic operations on the observed values. The commonly used descriptive measures fall in two general categories: measures of central tendency or location and measures of dispersion, scatter, or variability.

In most data sets, there is a tendency for values to cluster. A **measure of central tendency** locates the “middle” of the cluster. Thus, measures of central tendency are also called **measures of location**. Alternative measures of central tendency differ in the definition of the middle of the cluster.

1.2.1 Mean (Arithmetic Mean)

The mean is the statistic most commonly used to describe central tendency, and it is calculated from observed values on at least an interval scale of measurement. Batting average, grade-point average, and average temperature are examples of means. The word **average** refers to the **arithmetic mean** or to the centroid of a sample, typically referred to as the **mean**.

For a set of n observed values, x_1, x_2, \dots, x_n , the mean is the sum of the values divided by the number of values. The symbol denoting a mean is a line or bar over the symbol denoting the observed value. For example, when the observed values are denoted by x_i , $i = 1, 2, \dots, n$, the mean is denoted by \bar{x} and is pronounced “x-bar.” The mean is mathematically expressed as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.1)$$

The mean \bar{x} is in the same units as the observed values x_i that are summed to compute the mean.

Sometimes the values are grouped into intervals, and the number of values falling into each interval is provided. Techniques for obtaining groupings are presented in Section 1.4.3; such a grouping is one way of summarizing a sample of observed values. For each interval, a single number summarizes the values in the grouping; this number is called the group or class **mark**. Typically, this value is the midpoint of the group interval. Let v_j denote the class mark and f_j the frequency in the j^{th} group. If each x_i falling in the j^{th} group had the value of its class mark v_j , then the sum of observations in the j^{th} group would be exactly $f_j v_j$. Because all values of x_i are not actually equal to v_j but are distributed in the interval, the sum of the values in the j^{th} group is approximated by $f_j v_j$, and the sum of the observations in all groups is approximated by summing the $f_j v_j$ over the index j . The total number of observed values is the sum of the f_j . Thus, if k is the number of groups, the approximate mean calculated from grouped data is given by

$$\bar{x}_g = \frac{\sum_{j=1}^k f_j v_j}{\sum_{j=1}^k f_j} \quad (1.2)$$

1.2.2 Weighted Mean

In calculating the mean of a sample of values, it is assumed that all values are of equal importance. In situations where the values are not equally important, each value may be assigned a weight proportional to its relative importance, and a **weighted mean** is calculated. Let x_i , $i = 1, 2, \dots, n$, be observed values, and let w_i be the corresponding weights or relative importance of the observed values. The weighted mean is

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (1.3)$$

An example of a weighted mean is a grade-point average where the weights are the number of credit hours associated with each grade.

Example 1.1 Suppose that a sample of scrap is sent for analysis to two laboratories with the following results. A measure of analysis precision is reported as a quantity called the variance, defined in Section 1.3.2. The unweighted average of the two means is 7.40.

Laboratory	Mean	Variance
A	7.1 pCi/g	3.2
B	7.7 pCi/g	1.8

To find the weighted mean of the laboratories, weighted by the precision of each mean, the appropriate weight is the inverse of the variance. Using Equation 1.3, the mean weighted by the inverse of the variance is

$$\bar{x}_w = \frac{7.1/3.2 + 7.7/1.8}{1/3.2 + 1/1.8} = 7.48$$

Another use of weighted means is to estimate the overall mean with unequal-sized samples. If all observed values in the different samples are of equal importance, the mean of each sample is weighted by the sample

size. To combine k samples, where the j^{th} sample is of size n_j with mean \bar{x}_j , the overall mean is

$$\bar{x}_w = \frac{\sum_{j=1}^k n_j \bar{x}_j}{\sum_{j=1}^k n_j} \quad (1.4)$$

Note that this is similar to taking the mean of grouped data, where the \bar{x}_j serves as the mark and the n_j as the frequency. In this case, however, $n_j \bar{x}_j$ is the exact total for the j^{th} group rather than an approximation.

Example 1.2 Suppose that an operator in a nuclear fuel facility wants to know the average yearly number of fuel assemblies that have been shipped during the last 5 years. Available data show that for the past 5 years the average number of 24-rod assemblies shipped has been 227 per year. Also, during the last 2 years the average number of 27-rod assemblies shipped per year was 88. That is, for 3 of the 5 years, only 24-rod assemblies were shipped; for 2 of the 5 years, both 24-rod and 27-rod assemblies were shipped. The average number of assemblies shipped in the first 3 years is 227 per year and in the last 2 years is $227 + 88 = 315$ per year. From Equation 1.4, the overall average yearly number of assemblies shipped is

$$\bar{x} = \frac{3(227) + 2(227 + 88)}{3 + 2} = 262.2$$

1.2.3 Median

The **median** is a number that is greater than or equal to half of the observed values in a sample and simultaneously is less than or equal to the other half. Before the median can be determined, the observed values must be ordered. Thus, the measurements must be on at least the ordinal scale of measurement. The median, unlike the mean, is not affected by a few extremely large or small values.

Suppose that there are n observed values. These values are ranked in increasing order of magnitude (decreasing order works equally well). If n is odd, the median is the value in the middle position in the ordering. In the ordered series 2, 3, 3, 8, 11, 14, 19, the median is 8, because there are three values smaller than 8 and three values larger. If n is even, there is no

single middle position, so the median is defined as the average of the two values occupying the middle two positions. In the ordered series 2, 3, 3, 8, 11, 14, the two middle values are 3 and 8, and the median is $(3 + 8)/2 = 5.5$.

1.2.4 Mode

The **mode** is the value that occurs most frequently. Note that there can be more than one mode in a set of data. The mode can be found for measurements on any of the scales of measurement. When data are grouped, the group with maximum frequency is called the modal group. For grouped data measured on an interval or ratio scale, the mode is the mark of the modal group. Note that there can be more than one modal group. In the series 2, 3, 3, 8, 11, 14, the mode is 3 because 3 occurs twice and the other values each occur only once. In the series 2, 3, 3, 8, 8, 11, 14, the modes are 3 and 8.

1.2.5 Geometric Mean

For a set of n positive numbers x_1, x_2, \dots, x_n on at least an interval scale of measurement, the **geometric mean**, G , is defined as the n^{th} root of the product of the values; that is,

$$G = \left(\prod_{i=1}^n x_i \right)^{1/n} \quad (1.5)$$

By taking logarithms, this can be expressed as

$$\log G = \frac{1}{n} \sum_{i=1}^n \log x_i \quad (1.6)$$

That is, the geometric mean of the x_i values is the antilog of the arithmetic mean of the logarithms of the x_i values. The geometric mean is used for data that come from a lognormal distribution (discussed in Section 2.5.3) or for data with a multiplicative rather than an additive structure.

1.3 DESCRIPTIVE STATISTICS: MEASURES OF VARIABILITY

Measures of central tendency describe the middle of the data, whereas **measures of variability** describe the scatter or spread in a set of data. In a sample of measurements, the numerical values usually are not identical, but are dispersed due to random fluctuations in the measuring apparatus

and due to inherent differences in the items being measured. Measures of variability describe the spread of the values in a set of data and are also called **measures of dispersion** or **measures of scatter**. Both a measure of location (central tendency) and a measure of variability are needed to characterize a sample.

1.3.1 Range

The **range** is the difference between the largest and smallest values in a sample and is a good statistic for describing the spread in a set of data. The range is easily calculated from the observed values in a sample where measurements are on at least an interval scale. For small samples (less than 5 to 10 values), the range may be as stable as other measures of variability (i.e., in repeated samples of this size, the range often exhibits no more discrepancy from sample to sample than other measures of dispersion). One common, but somewhat invalid, objection to the range is that not all of the data are used. Because the range is computed from only the largest and smallest values, however, these cannot be determined unless all values are considered.

1.3.2 Variance and Standard Deviation

The most commonly used measure of variability is the **sample variance**, denoted by s^2 . It is calculated from measurements on at least an interval scale and is defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.7)$$

The divisor $(n-1)$ is used rather than n so that the sample variance is an unbiased estimator of the population variance (Section 3.4.1). The variance is related to parameters of many commonly used statistical distributions and is algebraically easy to manipulate for inferential purposes. Because the variance is computed using the square of deviations from the mean, it is heavily influenced by large deviations from the mean. All values, however, influence the variance. The sign of each deviation is not important because squaring results in all positive terms. It also results in units that are the square of the units of the observed values.

A convenient computational formula for s^2 , which is algebraically equivalent to the definition given in Equation 1.7, is

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \quad (1.8)$$

Another algebraically equivalent form is

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \quad (1.9)$$

Equations 1.8 and 1.9 are called **computing formulas** because they simplify hand computations. By applying Equations 1.1 and 1.8 or 1.9, the mean and variance are calculated by passing through the data once, accumulating the sum of values (Σx_i) and the sum of squared values (Σx_i^2). If the variance is computed directly from Equation 1.7, the data must be processed twice, first to compute the mean and then to compute the sum of squared deviations from the mean. Equations 1.8 and 1.9, however, should not be used on a computer, especially with large data sets, because serious roundoff errors can result in inaccurate calculation of the sample variance.

The variance is calculated from grouped data by applying the formula

$$s_g^2 = \frac{1}{n-1} \sum f_j (v_j - \bar{x}_g)^2 \quad (1.10)$$

where v_j is the class mark, f_j is the frequency of the j^{th} group, and \bar{x}_g is the mean estimated from Equation 1.2. This gives an approximation to the variance. Sheppard's correction, discussed by Kendall and Stuart (1973), can be applied to improve the approximation.

The **standard deviation** is defined as the positive square root of the variance. The units of the standard deviation are the same as the units of observed values and the measures of central tendency.

Another commonly used measure of variability is the standard deviation of a parameter estimate, which is often called the **standard error** of the parameter estimate.

1.3.3 Geometric Standard Deviation

The geometric mean, given by Equation 1.5, is the appropriate measure of central tendency to use for lognormally distributed data from a measurement scale that is at least an interval scale. The **geometric standard deviation (GSD)** is the corresponding measure of dispersion in such data. This statistic is the antilog of the standard deviation of the logarithms of the values. This definition is algebraically equal to

$$\text{GSD} = \exp \left\{ \frac{1}{n-1} \sum_{i=1}^n [\ln(x_i/G)]^2 \right\}^{1/2} \quad (1.11)$$

where G is the geometric mean, given by Equation 1.5.

1.3.4 Percentiles

Percentiles are summary values of data on at least an ordinal scale of measurement that can be used to convey information on either central tendency or dispersion. The $100p^{\text{th}}$ percentile of a sample is a value that exceeds at least $100p\%$ of the sample and is exceeded by at most $100(1-p)\%$. Sample percentiles are easily determined if data are ranked or ordered. In Section 1.2.3, the median is defined as the 50th percentile. The 25th and 75th percentiles have special names, the first (or lower) and third (or upper) **quartiles** of the data. The first quartile is the value that equals or exceeds one-quarter of the data and also is equaled or exceeded by three-quarters of the data. Similarly **deciles** are the 10th percentile, 20th percentile, and so on. Quantities such as quartiles, deciles, and percentiles are often called **quantiles**, and functions of quantiles are used as measures of dispersion. The most commonly used of these is the **interquartile range**, defined as the difference between the upper and lower quartiles.

1.4 DATA DISPLAY: SINGLE VARIABLE

It is often difficult to see patterns in tables of numbers. Thus, displays that summarize data can be used to convey the information in the data at a glance, based on the use of fewer numbers. Characteristics and patterns that are not readily apparent in a collection of numbers often can be easily determined from appropriate summaries or displays of the data. Display techniques differ in the scales of measurement that are valid and in the number of variables that are displayed simultaneously.

1.4.1 Pie Charts

One common type of diagram is the **pie chart**. A “pie” or circle representing the whole is cut into slices or sectors corresponding in area to proportions that comprise the whole. The pie chart can be used to display

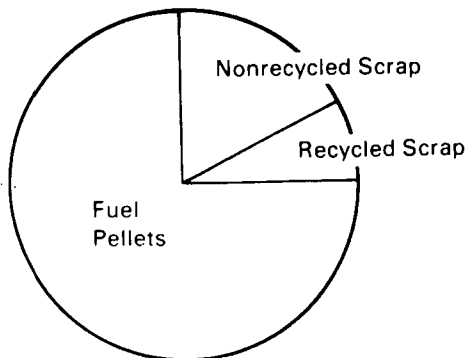


Figure 1.1 Pie chart illustrating disposition of fuel pellets.

data from any measurement scale. For a scale of measurement higher than nominal, however, information is lost if the data are reduced to a nominal level for a pie chart display. Figure 1.1 is an example of a pie chart that displays the proportions of fuel pellets with the indicated dispositions.

1.4.2 Bar Charts

A **bar chart** is used to display nominal or ordinal scale measurements. Each bar represents one group, and the bar height corresponds to that group frequency. Bars are of equal width, so that area and height of the bars are proportional. Bars may be oriented horizontally or vertically. If the categories are on an ordinal scale, the bars are placed in sequence. Figure 1.2 displays the monthly number of fuel rod rejects at a fuel fabrication facility during a particular year.

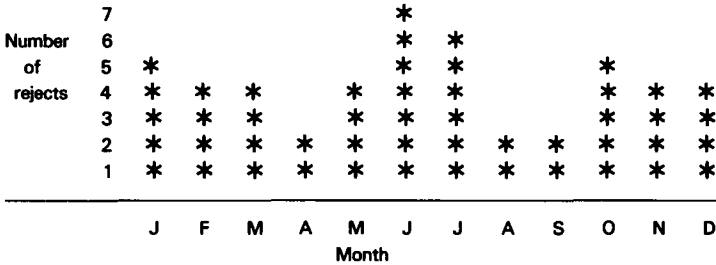


Figure 1.2 Bar chart illustrating monthly rejects of fuel rods.

1.4.3 Histograms

A **histogram** is similar to a bar chart and is used for displaying frequencies of values of a variable measured on an interval or ratio scale. The heights of the bars represent frequencies. Each bar corresponds to a value of a discrete variable or an interval of values of a continuous variable. Bar width represents interval size. In the following discussion, intervals of equal width are assumed, so bar height and area are proportional.

Consider Table 1.1 which displays the weights of six samples of 25 uranium ingots for input to a nuclear fuel production cycle. From visual examination of Table 1.1, it is difficult to see the central tendency and dispersion of ingot weights in the samples. Thus, a histogram is appropriate to summarize the central tendency and dispersion of the observed values of ingot weight.

The first step in constructing a histogram is to group the data. A grouping is chosen to condense information for the specific display purpose. A typical grouping is based on intervals of equal width. A grouping is successful if it conveys significant information from the data but does not display unnecessary detail. Sometimes a grouping may not condense the data enough, and the histogram may not convey the central tendency

TABLE 1.1
Six Samples of Size 25 of Uranium Ingot Weights†

Sample Number					
1	2	3	4	5	6
425.0	426.7	423.3	429.4	427.9	422.1
422.5	424.4	427.8	428.9	425.3	427.3
424.9	422.4	426.1	424.8	432.4	419.7
431.7	432.2	422.4	427.3	427.3	425.7
426.3	427.8	424.8	428.0	426.3	425.9
424.5	431.3	431.2	427.3	418.5	431.6
426.1	425.8	429.8	429.5	425.3	424.6
423.1	426.8	430.9	423.9	421.9	421.8
428.3	424.8	427.0	425.1	425.2	432.3
423.2	423.6	427.9	427.9	428.5	428.8
428.2	424.8	421.0	423.6	428.0	425.7
429.1	429.7	419.6	421.3	426.8	425.2
424.2	430.3	424.6	430.0	423.5	430.0
429.7	423.2	428.8	425.4	427.5	424.9
424.8	431.0	427.9	423.6	421.7	426.6
427.2	428.0	428.0	429.7	427.4	426.2
428.3	426.6	428.4	427.1	427.5	426.2
429.3	425.4	423.1	426.9	425.7	434.2
421.8	427.3	425.2	427.3	425.7	420.4
424.0	426.0	424.9	430.5	426.3	424.9
428.0	423.3	431.1	426.4	429.0	423.3
426.5	426.3	429.9	427.4	424.2	428.2
427.2	429.2	425.9	426.6	425.5	419.2
424.5	425.1	424.4	424.7	427.7	421.2
425.9	425.4	427.9	423.7	428.6	428.8

†In kilograms.

and dispersion because there are too few observations in each interval. If this is the case, fewer groups based on wider intervals should be used. On the other hand, if the shape of the distribution of observed values is not apparent from the histogram because all observations are in a few intervals, then the number of groups should be increased using narrower intervals. The samples in Table 1.1 are summarized in Figure 1.3 in histograms with intervals of unit width in which one less digit is displayed in the figure than in Table 1.1.

Let x be the observed weight. The group intervals in Figure 1.3 are defined by the inequalities $417.5 < x \leq 418.5$, $418.5 < x \leq 419.5$, $419.5 < x \leq 420.5$, and so on. Because of the $<$ and \leq signs, there is no ambiguity as to which class each weight belongs. The categories can be

defined instead as 417.51 through 418.50, 418.51 through 419.50, 419.51 through 420.50, and so on. Here, ambiguity is avoided by using one additional significant digit to define the interval boundaries. In both cases, the middles of the intervals are the integers 418 through 434, and one fewer digit is displayed than is recorded in Table 1.1.

Based on this grouping for the six samples, the six histograms in Figure 1.3 are generated. These histograms provide a much clearer picture of the central tendency and dispersion within each sample than the columns of numbers in Table 1.1. Note, however, that the central tendency and especially the dispersion of the values appear to vary among samples.

Middle of interval	Frequency					
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
418	0	0	0	0	1*	0
419	0	0	0	0	0	1*
420	0	0	1*	0	0	2**
421	0	0	1*	1*	0	1*
422	2**	1*	1*	0	2**	2**
423	2**	2**	2**	0	1*	1*
424	4****	2**	1*	4****	1*	0
425	3***	5*****	4****	4****	4****	4****
426	4****	3***	2**	1*	4****	5*****
427	2**	4****	1*	7*****	5*****	2**
428	4****	2**	6*****	2**	4****	1*
429	2**	1*	1*	3***	2**	2**
430	1*	2**	2**	3***	0	1*
431	0	2**	3***	0	0	0
432	1*	1*	0	0	1*	2**
433	0	0	0	0	0	0
434	0	0	0	0	0	1*

Figure 1.3 Histograms of uranium ingot weights by sample number from Table 1.1.

Assume that the six samples are known to be from the same population and, thus, can be combined. Methods for testing this assumption are given in Chapter 5. If the six samples are combined and a histogram is drawn as in Figure 1.4, it appears that the central tendency in the data is somewhere between 425 and 428 kg, with the values ranging from 418 to 434 kg. The distribution of weights is approximately symmetrical. With the combined sample, the shape of the histogram gives a better indication of the distribution of observed values than the individual histograms in Figure 1.3. This illustrates the increased information that can be gained by increasing the sample size.

Middle of interval	Frequency
418	1*
419	1*
420	3***
421	3***
422	8*****
423	8*****
424	12*****
425	24*****
426	19*****
427	21*****
428	19*****
429	11*****
430	9*****
431	5*****
432	5*****
433	0
434	1*

Figure 1.4 Combined histogram of ingot weights from Table 1.1.

1.4.4 Stem-and-Leaf Displays

A histogram uses data measured on an interval or ratio scale, grouped into intervals to reduce data to an ordinal level for display. This reduction of measurement scale discards some of the information in the data. A **stem-and-leaf display** retains all or almost all of the information in the interval or ratio scale measurements. It shows the range of values in the data, the gaps where no values occur, the concentration or clustering of values, the symmetry of values, values that are markedly different than the rest, and any unusual patterns in the data. These are features that may go unnoticed in a table or listing of data.

One feature of a stem-and-leaf display is the illustration of most of the digits of the observed values (some information is lost with truncation). Each value is split into two parts, leading digits (stem) and trailing digits (leaf). Usually, trailing digits are further split into displayed and discarded digits. For example, the number 62367 might be split into 62 as leading digits and 367 as trailing digits. Of trailing digits, the 3 would be displayed and the 67 discarded. With numbers that have only two or three digits, no digits are discarded.

In a stem-and-leaf display each unique combination of leading digits is displayed on a separate line in increasing order. Any missing combinations of leading digits in the counting sequence are also displayed on a separate line. The retained trailing digits for all values with the same leading digits are displayed on the same line as the leading digits. For the number 62367, if 6 and 2 (displayed as 62) are chosen as the leading digits and 3

as the displayed trailing digit, the 62 is displayed on a line. On the same line the digit 3 is displayed, along with the third digit of any other five digit values in the data set between 62000 and 62999. The displayed trailing digits are arranged in increasing order of magnitude for each stem value.

Example 1.3 Construct a stem-and-leaf display of the five numbers 493, 506, 478, 497, and 502, with the first two digits as the leading digits (stem) and the third digit as the trailing digit (leaf). The stem-and-leaf display is

Stem	Leaf
47	8
48	
49	37
50	26

The stem digits are to the left and are not repeated for each observed value. The leaf digits are on the right, ordered by magnitude, and on the row corresponding to the two digits in their stem. The stem value of 48 is included with a blank leaf space to indicate that no values occur between 480 and 489.

Choice of the position of the stem-and-leaf split is determined by the range and number of values to be displayed. For large samples with four or more digits per value, a three-digit stem is typically used if the range of stem values is not large. Because decimal point location is lost when a value is split into a stem and leaf, the display notes where the decimal point falls; this is usually provided in the heading. A stem-and-leaf display of the data in Table 1.2 is given in Figure 1.5.

In the heading of the stem-and-leaf display in Figure 1.5, the placement of the decimal point is indicated in two ways. The first line indicates that the leaf digit represents the units part of a number. The second line, with the space between the first 266 and 2 (representing the space between stem-and-leaf columns), shows that a stem of 266 and a leaf of 2 represents 2662 kg.

A feature not apparent in Table 1.2 is emphasized by Figure 1.5; i.e., all leaf digits are even numbers. This suggests that measurements were recorded to the nearest even kilogram. Such patterns are lost in a histogram in which each of the leaf digits is replaced by an asterisk.

TABLE 1.2

Uranium Hexafluoride Cylinder Weights†‡

2676	2684	2690	2682	2684
2674	2690	2686	2660	2696
2682	2710	2692	2658	2686
2686	2670	2690	2686	2694
2690	2690	2682	2672	2708
2672	2680	2680	2664	2684
2640	2672	2678	2682	2682
2682	2692	2690	2696	2660
2672	2688	2666	2666	2674
2692	2662	2642	2676	2686

†In kilograms.

‡Simulated data with a mean of 2680 kg.

Leaf Digit Unit = 1
266 2 Represents 2662

Stem	Leaf
264	02
265	8
266	002466
267	0222244668
268	00222222444666668
269	000000222466
270	8
271	0

Figure 1.5 Stem-and-leaf display of cylinder weights from Table 1.2

1.4.5 Boxplots

Frequently, less detail than a stem-and-leaf display is sufficient. A display of all the information may be confusing, but an overall summary such as a mean and standard deviation is too brief. In such cases, **boxplots** may be appropriate. Boxplots require that the observed values be interval or ratio scale measurements.

Boxplots use a median to describe the central tendency of data. To represent the spread, each half of the data is split in half again by the **hinges**. Hinges are found in the same way that the median was found in Section 1.2.3; the data are split into two parts at the median, and the hinges are the medians of each of the two parts. Depending on the num-

ber of values in the data, the hinges are either equal to or approximately equal to the upper and lower quartiles. A boxplot summarizes data by using five numbers: the median, the two hinges, and the maximum and minimum values. A boxplot is made by drawing a box or rectangle between the hinges. A vertical line is placed across the box at the median. A horizontal line is extended from each hinge to the maximum and minimum values; these lines are called **whiskers**.

A set of data on shipper–receiver differences is displayed in Table 1.3. Assume that the time between shipper and receiver weighings is relatively short and that the shipments are made monthly at approximately the same time, so that the time intervals are approximately equal. Each observation consists of three values (t_i, s_i, r_i) , where i is the index ($i = 1, \dots, 10$), t_i is month of shipment, and s_i and r_i are shipper and receiver net weight values. Because the values of index and month are the same, only one is needed. Thus, the shipper–receiver data have up to five variables at various scales of measurement: cylinder index (nominal or ordinal scale), time (date) of shipper weighing (interval scale), time (date) of receiver weighing (interval scale), shipper net weight (ratio scale), and receiver net weight (ratio scale). In addition, in Table 1.3 the difference in net weight between shipper and receiver values is calculated.

Figure 1.6 is a boxplot of the shipper–receiver differences $(s_i - r_i)$ from Table 1.3. An overall impression is displayed, including spread and symmetry. The median is 1.34 kg, the spread between hinges is 1.62 kg (0.22 to 1.84 kg), and the range is 5.52 kg (-2.06 to 3.46 kg). The distribution is asymmetrical because the median is not in the middle of the box. Because the entire box is above zero, there is a good indication that the shipper weights tend to be larger than the receiver weights.

TABLE 1.3
Net Weight of UF₆ Cylinders by Month†

Index, i	Month, t_i	Shipper, s_i	Receiver, r_i	Difference, $s_i - r_i$
1	1	1471.22	1468.12	3.10
2	2	1470.98	1469.52	1.46
3	3	1470.82	1469.22	1.60
4	4	1470.46	1469.26	1.20
5	5	1469.42	1465.96	3.46
6	6	1468.98	1470.80	-1.82
7	7	1469.10	1467.89	1.21
8	8	1470.22	1472.28	-2.06
9	9	1470.86	1469.02	1.84
10	10	1470.38	1470.16	0.22

†In kilograms.

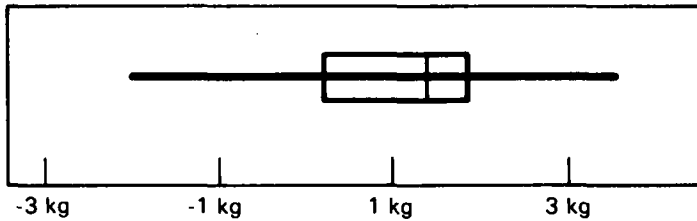


Figure 1.6 Boxplot of shipper-receiver differences ($s_i - r_i$) from Table 1.3.

Samples can be compared visually by placing several boxplots in the same figure. For example, to further compare the shipper and receiver data, the corresponding boxplots are placed together in Figure 1.7. The shipper and receiver sample medians do not coincide (the shipper weights tend to be larger than the receiver weights), and the shipper measurements are more precise (less spread out) than the receiver measurements.

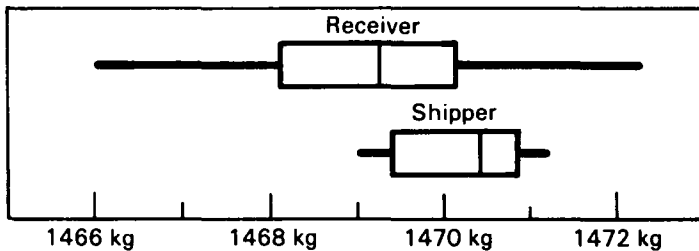


Figure 1.7 Boxplots of shipper and receiver weights from Table 1.3.

1.5 DATA DISPLAY: MULTIPLE VARIABLES

The shipper-receiver data displayed in Figure 1.7 form ordered pairs. That is, for each value reported by the shipper, there is a corresponding value reported by the receiver. The pairing occurs because the same uranium hexafluoride cylinder is measured. An ordered pair of values is indicated by enclosing the values in parentheses: if x is used to indicate one variable and y the other, then (x,y) is an ordered pair. For the shipper-receiver data, values are referred to by s and r , so the ordered pair is represented by (s,r) or (r,s) . In addition, the observations in this example are time-ordered, because the time sequence in which cylinders were shipped and weighed is known.

For most months, the shipper values are larger than the receiver values. If only randomness were influencing the data, about five positive and five

negative differences would be expected. In this case, two negative and eight positive differences are observed. This suggests that, on the average, shipper values may be larger than receiver values due to some nonrandom influence. Based on the Sign Test (Chapter 9) for 10 observed differences, eight or more positive and two or fewer negative differences are expected to occur by chance about 5.4% of the time in situations where there is no true difference. Thus, there is evidence of nonrandom structure in the data that may be further explored graphically. In the following sections, different graphs of the shipper–receiver data are presented.

1.5.1 Scatterplots

Pairs of observed values are typically displayed as **scatterplots** or **x–y plots** to show how x and y are related. Using the data in Table 1.3, if shipper values are plotted on the vertical axis (y -axis) or **ordinate** and receiver values on the horizontal axis (x -axis) or **abscissa**, each (r,s) pair is represented by a point on the graph. This plot is shown in Figure 1.8.

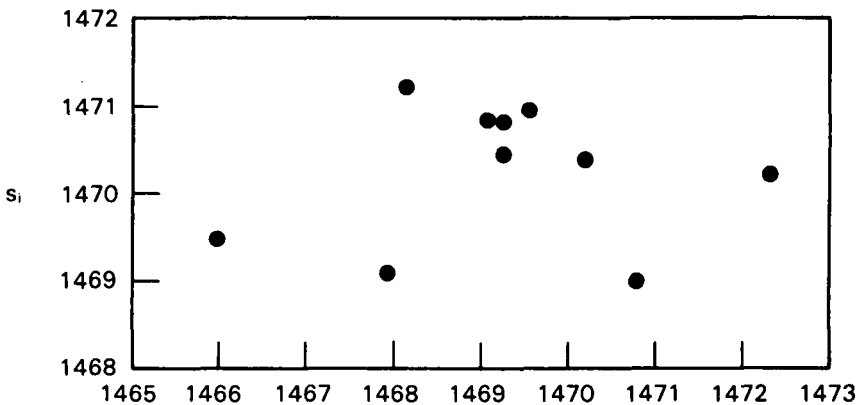


Figure 1.8 Scatterplot of pairs of shipper and receiver weights from Table 1.3.

The month of shipment can be included as a third variable in a scatterplot by replacing each plotted point with the index number for the month of shipment. Such a display is shown in Figure 1.9. The diagonal line indicates where the points should fall if the shipper weight equals the receiver weight.

The points in Figure 1.8 or 1.9 do not follow any simple pattern. However, during the months 5, 6, 7, and 8, the receiver weights show a greater variability than shipper weights and the other six receiver weights, but the central tendency of these four receiver weights is close to the yearly central tendency of all receiver weights.

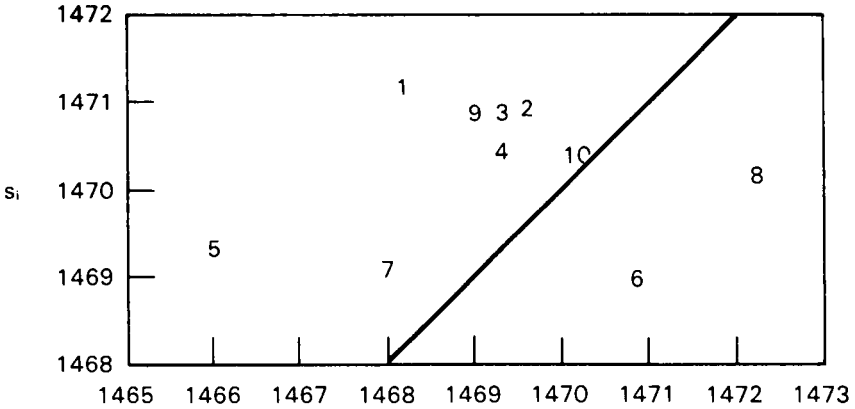


Figure 1.9 Scatterplot of pairs of shipper and receiver weights from Table 1.3 also displaying month of shipment.

1.5.2 Time Series Plots

The data in Figure 1.9 suggest that there may be a seasonal effect. Thus, it is of interest to plot time (month) on the x-axis. A plot of observed values vs. time is called a **time series plot**. Several such plots can

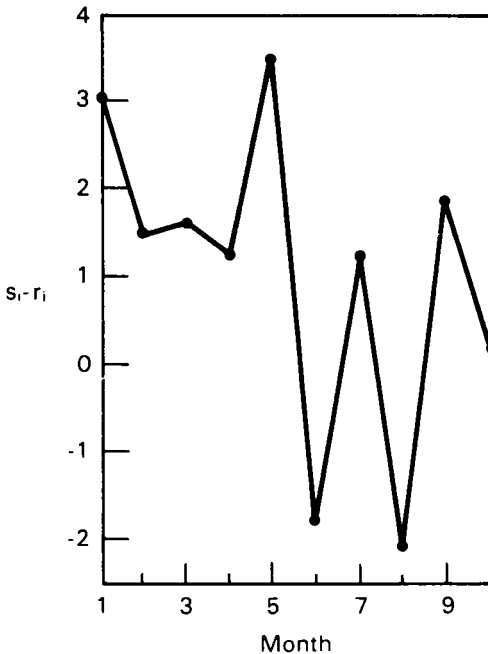


Figure 1.10 Time series plot of shipper-receiver differences from Table 1.3.

be derived from this example set of data. One plot that may be useful would have two lines on it: one describing the shipper's values and a second describing the receiver's values. Another useful plot is of the shipper-receiver differences, as shown in Figure 1.10.

Note that the clustering in Figures 1.8 and 1.9 is not apparent in Figure 1.10. There is a suggested trend from upper left to lower right in Figure 1.10, but there is also substantial variability. In Section 1.6, methods are presented for "filtering out" some of the variability so that underlying trends are more obvious.

1.6 DATA SMOOTHING

When x-values (abscissa) are equally spaced, it is useful for y-values (ordinate) to change smoothly from point to point to indicate trends in a set of data. The process of smoothing consists of filtering some of the randomness from the y-values. This section presents techniques for such filtering, where the discussion is restricted to equally spaced x-values. In this case, the y-values are considered to be a sequence of values. Methods also exist, however, for unequal spacing.

Smoothing is the operation of displaying the dominant trends in a sequence and ignoring some of the irregularities. Consecutive overlapping segments of the sequence are used for smoothing; for example, the first four values are used, then the second through the fifth, then the third through the sixth, and so on. Because the segments overlap, changes in the smoothed data are smoother than in the original data. A mathematical algorithm that removes variability is called a **filter**. The filters discussed in this section use medians and means to smooth a sequence.

Using more than one smoother or filter sequentially is called **compound smoothing**. Through a combination of smoothing procedures, the definition (or resolution) of an underlying trend is improved. Early steps in a compound smoother concentrate on moderating the effect of unusual values in the sequence. Later steps concentrate on creating a smooth trend.

A more complete discussion of smoothing is found in Chapters 7 and 16 of Tukey (1977). A review of smoothing algorithms is found in Velleman (1980).

Although several smoothing methods are presented, no rule is given as to which method is preferred. This is often a matter of individual choice.

1.6.1 Running Medians

Suppose that each observed value, except the end values in a sequence, is replaced by the median of itself and its two neighboring values. This is an example of a **3-point running median smoother**, so named because each value is replaced with a median. If the two neighboring values on each side of a value and the value itself are used to estimate the median, a **5-point**

running median smoother results, and so on. Methods of handling the first and last points in a 3-point smoother or the first and last two points in a 5-point smoother, and so on, are presented in Section 1.6.4. As a simplified procedure, the original first and last points are copied into the smoothed sequence. To illustrate running medians, a 3-point smoother is applied twice to the shipper–receiver difference data from Table 1.3. The results are displayed in Table 1.4. The column labeled “Residual” is calculated as the difference between the column labeled “Difference” and the column labeled “2nd Smoothing.”

TABLE 1.4

Three-Point Smoothing of Shipper-Receiver Differences from Table 1.3

Month	Difference		1 st smoothing		2 nd smoothing	Residual
1	3.10	----copy----	3.10	----copy----	3.10	0.0
2	1.46		1.60		1.60	-0.14
3	1.60		1.46		1.60	0.0
4	1.20		1.60		1.46	-0.26
5	3.46		1.20		1.21	2.25
6	-1.82		1.21		1.20	-3.02
7	1.21		-1.82		1.21	0.0
8	-2.06		1.21		0.22	-2.28
9	1.84		0.22		0.22	1.62
10	0.22	----copy----	0.22	----copy----	0.22	0.0

Figure 1.11 shows the effect of the first 3-point median smoothing; a smooth curve is obtained except for the 7th month. Figure 1.12 illustrates that applying the second 3-point median smoother aligns this value. For these data, the third 3-point median smoother is virtually identical to the second.

When smoothing with an even number of values (an even segment), the x -values are used as well as the y -values. A description of a **4-point running median smoother** illustrates the principle. Recall that the median of four values is obtained by ordering the values and then averaging the middle two values.

When using a 4-point median smoother, it is unclear which of the two central x -values should be associated with the median y -value. The median of an odd length smoothing segment is associated with the middle x -value of the segment. For example, with a 3-point smoothing segment, the median y -value from the data segment defined by (x_{j-1}, y_{j-1}) , (x_j, y_j) , (x_{j+1}, y_{j+1}) is associated with the abscissa position x_j . The center of an even length segment, however, is between two x -values. Thus, the new y -values are aligned with the original x -values by averaging the two 4-point

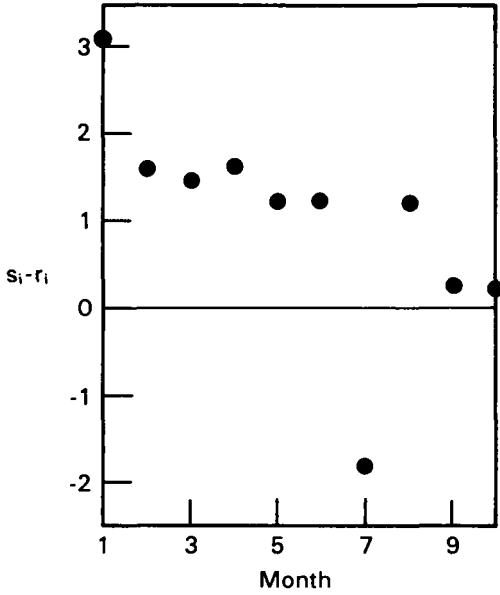


Figure 1.11 First smoothing by 3-point running median of shipper-receiver differences from Table 1.4.

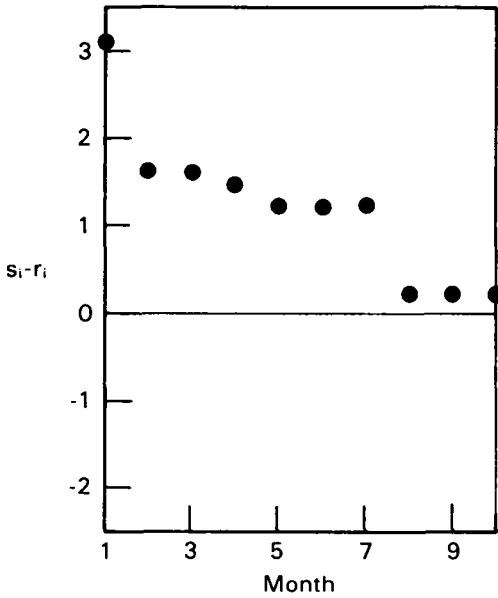


Figure 1.12 Second smoothing by 3-point running median of shipper-receiver differences from Table 1.4.

running medians on either side of the two middle x -values. This is called **recentering**. The operations are as follows:

x-values	. . .	5	6	7	. . .
y-values	. . .	y_5	y_6	y_7	. . .
smoothed by 4's	$Z_{4.5}$	$Z_{5.5}$	$Z_{6.5}$	$Z_{7.5}$	
recentered		Z_5	Z_6	Z_7	

Note that recentering is a 2-point running median of the 4-point running medians. Algebraically, a 4-point recentered running median replaces each y_j with

$$Z_j = 1/2 [\text{median}(y_{j-2}, y_{j-1}, y_j, y_{j+1}) + \text{median}(y_{j-1}, y_j, y_{j+1}, y_{j+2})] \quad (1.12)$$

For illustration, the 4-point recentered smoother is applied to the shipper-receiver differences from Table 1.3. Table 1.5 summarizes the computations.

TABLE 1.5

Recentered 4-Point Running Median Smoothing of Data From Table 1.3

Month	Difference	Smoothed	Residual
1	3.10	-----copy-----	3.10
2	1.46	-----copy-----	1.46
3	1.60	1.53	1.53
4	1.20	1.53	1.47
5	3.46	1.40	1.31
6	-1.82	1.21	0.45
7	1.21	-0.31	-0.31
8	-2.06	-0.31	0.21
9	1.84	0.72	1.84
10	0.22	-----copy-----	0.22

Figure 1.13 is a plot of the recentered data. A comparison with Figure 1.12 shows that a double 3-point and a 4-point median smoother have about the same effect on the first half of the sequence but that the 4-point smoother does not completely filter the randomness in the last half of the sequence.

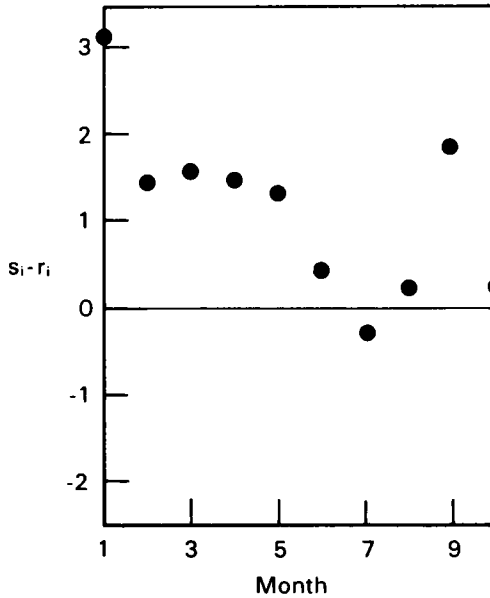


Figure 1.13 Four-point running median smoothing and recentering of shipper-receiver differences from Table 1.5.

1.6.2 Running Averages

A filter that has less of a smoothing effect than the median filters may be preferred. One possibility is a **running average** which smooths sequences by replacing each value with the average of itself and neighboring values. Often a **running weighted average** is used, in which neighboring values closest to the value being replaced are given the highest weights. For example, y_j may be replaced by $\frac{1}{4} y_{j-1} + \frac{1}{2} y_j + \frac{1}{4} y_{j+1}$. This is a weighted average because the coefficients ($\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$) are unequal and add to one. Although an unlimited number of running averages and running weighted averages are possible, this particular one is useful in application and is called **hanning**. Any running weighted average is affected by even a single unusual value, so such smoothers are often used only after unusual values are filtered by a median smoother.

1.6.3 Smoothing the Endpoints

Endpoints in a sequence cannot be smoothed in the same way as values in the middle because the endpoints are not surrounded by values on both sides. With a 5-point smoother, the problem is alleviated by substituting a 3-point smoother near the ends of the sequence. Likewise, a 2-point

smoother can be substituted for a 4-point smoother near the ends. This substitution smooths all values except the first and last points. Smoothed values for the first and last points are obtained by extrapolating from the smoothed curve to the endpoints.

One approach is to find the line that passes through the two smoothed values adjacent to the endpoints, then to extrapolate this line to the endpoints. For equally spaced values on the x-axis, the distance or time between values is denoted as Δx . The y-values (y_1, y_2, \dots, y_n) are smoothed into (Z_2, Z_3, \dots, Z_{n-1}). Linear extrapolation obtains values for Z_1 and Z_n . The slope of the extrapolation line at the beginning of the sequence is

$$(Z_3 - Z_2)/\Delta x \quad (1.13)$$

and the estimated first value is

$$Z_1 = Z_2 - \Delta x(Z_3 - Z_2)/\Delta x = 2Z_2 - Z_3 \quad (1.14)$$

The estimate for the final value is

$$Z_n = 2Z_{n-1} - Z_{n-2} \quad (1.15)$$

When using compound smoothers, it is unnecessary to perform end-point extrapolation between each step. Extrapolation should be performed, however, both before and after hanning.

1.6.4 Examining the Residuals

Studying a smoothed sequence is part of examining a time series. It is also important to study **residuals**, which are the differences between filtered and original values. Examining residuals can reveal **outliers**, patterns or trends in the sequence that are short term. Residuals can also reveal portions of the sequence subject to more variability than other portions. Figure 1.14 displays the residuals for the shipper–receiver differences example in which the double 3-point smoother was applied.

No rules for examining residuals are formulated because the purpose of such an examination is to find unusual or unexpected trends in the sequence. Sometimes it is helpful to smooth residuals. This can yield a secondary trend. What is to be done after smoothing has identified a trend is not discussed here. Working with an algebraic definition of a trend is discussed in Chapter 6.

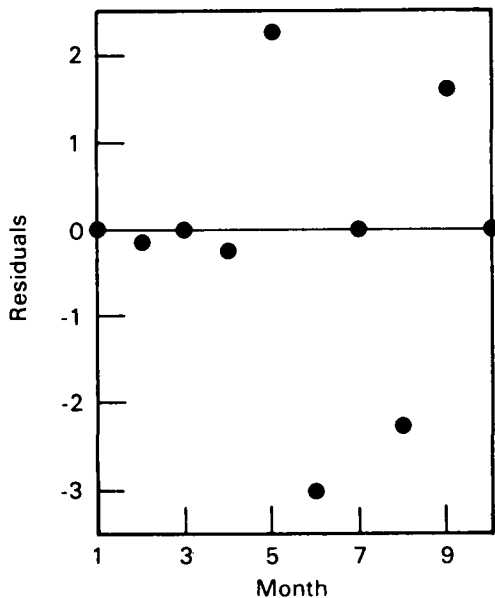


Figure 1.14 Residuals after smoothing of shipper-receiver differences with the double 3-point median smoother from Table 1.4.

REFERENCES

- Kendall, M. G., and A. Stuart, 1973, *The Advanced Theory of Statistics*, Hafner Publishing Co., New York.
- Tukey, J. W., 1977, *Exploratory Data Analysis*, Addison-Wesley Publishing Co., Reading, Massachusetts.
- Velleman, P. F., 1980, Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms, *J. Am. Stat. Assoc.*, 75: 609-615.
- , and D. C. Hoaglin, 1981, *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press, Boston.

CHAPTER 2

Probability

2.0 INTRODUCTION

Statistics enters into scientific investigation through experimentation and observation. Scientific investigation is a means for acquiring information from which a conclusion can be drawn. Most statements resulting from scientific investigations are only inferences which are uncertain in character. In many instances, part of the uncertainty is random in nature and can be described using probability theory. The drawing and evaluation of inferences in the face of random uncertainty is the essence of statistics.

In this chapter some basic probability concepts and formal definitions are provided that are fundamental to the development and application of the statistical methods presented in subsequent chapters. This is not meant to be a comprehensive presentation of probability theory. The reader interested in a more thorough treatment of the subject can find varying levels of rigor in such texts as Feller (1968), Hines and Montgomery (1980), Hogg and Craig (1978), and Rohatgi (1984).

2.1 BASIC PROBABILITY CONCEPTS

Prior to a formal discussion of probability, a few basic concepts and terms are introduced to facilitate the formal definitions and theorems[†] which are presented in this section. First, the concept of experimentation is introduced.

According to Webster's Ninth New Collegiate Dictionary (1983), an **experiment** is "an operation carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law." Similarly, the American Heritage Dictionary (1982) defines an experiment as "a test made to demonstrate a known truth, to examine the validity of a hypothesis, or to determine the efficacy of something previously untried." These definitions include the essence of the notion of an experiment as used in most statistics texts. For example, rolling dice or tossing a coin may be looked upon as experiments.

[†]The expression "theorems" is used somewhat loosely to denote various properties, propositions, theorems, etc., which result from the definitions. Although not strictly correct, this will reduce the number of terms to be absorbed by the reader.

Each possible result of an experiment is called an **outcome** of the experiment. The numbers that turn up on the rolled dice or the face observed on the tossed coin are outcomes of the corresponding experiments. Each of the outcomes that can occur in a single trial of an experiment is called an **elementary event** or a **sample point**. The collection of all possible sample points for a given experiment is called the **sample space**. An **event**, sometimes called a **chance event**, is a specified subset of the sample space (i.e., the sample points in this subset possess some specified characteristic).

As an example, suppose that an experiment consists of rolling a die. The possible outcomes are the numbers 1, 2, 3, 4, 5, 6. That is, the sample space has a total of six sample points. An event of interest might be that of observing an even number. This event includes the sample points 2, 4, 6, and is a subset of the sample space.

Probability is a measure of the likelihood of occurrence of a chance event. The specific approach to probability used here is an axiomatic one, in which the probability of an event is simply a number associated with that event, and where the probabilities of a collection of events are constrained to obey a certain set of axioms. Although other approaches to probability exist, the axiomatic approach is the most consistent with the subject matter of this book.

Before presenting the definition and axioms of probability, it is necessary to introduce a number of definitions and concepts from set theory.

Definition 2.1

A **set** is a collection of elements.

Definition 2.2

The **universal set** is the set consisting of all elements under consideration. The universal set is sometimes referred to as a **space**.

Definition 2.3

The **null set** is the set containing no elements.

Definition 2.4

Associated with each set A is another set A' called the **complement** of A . The set A' contains all elements of the universal set which are not elements of A .

Definition 2.5

The **union** of any two sets A and B is the set consisting of all elements which are either in A or in B or in both A and B . The union of A and B is commonly denoted by $A \cup B$ or by $A + B$.

Definition 2.6

The **intersection** of any two sets A and B is the set consisting of all elements which are simultaneously in both A and B . The intersection of A and B is commonly denoted by $A \cap B$ or simply AB .

To illustrate the concepts presented in Definitions 2.1 through 2.6, consider an ordinary deck of 52 playing cards. Let the universal set be the 52 cards. Define the set A to be the 13 cards belonging to the suit hearts. Then the set A' would be the 39 cards which are not hearts. Define the set B to be the 4 cards with a king for their "face value." The union of sets A and B is the set which includes all cards showing a heart, or a king, or both. The intersection of sets A and B is the set which includes all cards showing both a heart and a king (the king of hearts card is the only element in this set).

A useful device for illustrating the properties of the algebra of sets is the **Venn diagram**. In such a diagram, the points interior to a rectangle constitute the universal set. Arbitrary sets within the universal set (i.e., subsets of the universal set) are represented by points interior to circles within the rectangle. In the Venn diagram displayed in Figure 2.1, the set A is shaded by vertical lines, and the set B is shaded by horizontal lines. Because in this figure the set AB is not the null set, it appears as the crosshatched area. In the above card example, A includes all hearts, B includes all kings, the crosshatched area includes the king of hearts, and the area outside the two circles but inside the rectangle includes all other cards.

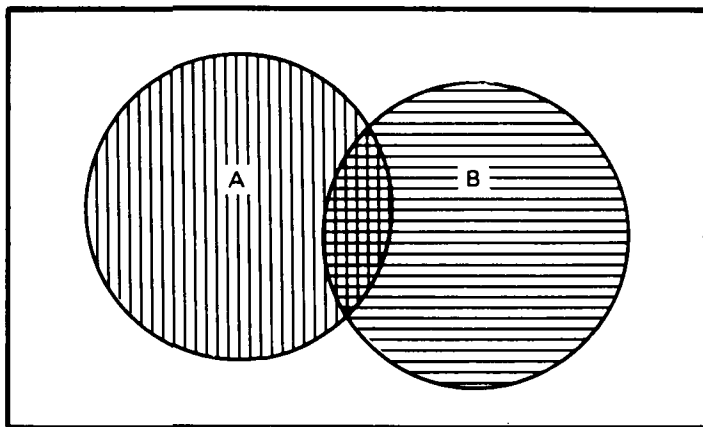


Figure 2.1 A Venn diagram.

Counting the number of elements in a set is an essential part of probability theory. The following concepts are useful.

Definition 2.7

The number of elements in an arbitrary set A is denoted by $n(A)$.

Theorem 2.1

For any two sets A and B , $n(A \cup B) = n(A) + n(B) - n(AB)$.

Theorem 2.2

If A and B are **mutually exclusive** (have no points in common), then AB is the null set. Thus, $n(AB) = 0$, and $n(A \cup B) = n(A) + n(B)$.

In the previous card example there are $n(A) = 13$ hearts and $n(B) = 4$ kings. The set AB includes only the king of hearts, thus $n(AB) = 1$. Then $n(A \cup B) = 13 + 4 - 1 = 16$. That is, there are 16 cards that show a heart, or a king, or both. Note that in computing $n(A \cup B)$, we first add $n(A)$ and $n(B)$, but because the king of hearts belongs to both A and B , it has been counted twice. Thus, $n(AB)$ is subtracted from the sum so that no elements are counted twice.

Permutations and combinations are concepts used for counting the different subgroups and arrangements that can be formed from a given set. A **permutation** is a particular sequence or arrangement of a given set or subset of elements. A **combination** is the set or subset without reference to the order of the contained elements. In the following definitions, the factorial notation $n!$ is used where $n! = n(n - 1)(n - 2) \dots (2)(1)$. The convention $0! = 1$ is assumed.

Definition 2.8

The number of **permutations** (distinct arrangements) which can be formed from n elements taken r at a time is $P(n,r)$. It can be shown that $P(n,r) = n!/(n - r)! = n(n - 1) \dots (n - r + 1)$.

Definition 2.9

The number of **combinations** (different subsets) which can be formed from n distinct elements taken r at a time is $C(n,r)$. It can be shown that $C(n,r) = P(n,r)/r! = n!/(n - r)!r!$.

To illustrate permutations and combinations, consider once again a deck of 52 playing cards. In being dealt a five-card poker hand, a player is interested only in the composition of the hand and not in the order in which the cards are dealt. Thus, the player might be interested in counting the number of possible five-card poker hands that could be dealt to him from a 52-card deck without regard to the arrangement of the cards in a given hand. This is simply the number of combinations which can be formed from 52 objects taken five at a time and is computed as $C(52,5) = 52!/47!5! = 2,598,960$.

Now suppose that for some reason the player is interested in counting the number of possible distinct five-card arrangements that could be dealt from the 52 cards. Conceptually, this is analogous to first being dealt one of the 2,598,960 possible five-card hands and then arranging the five cards within the hand. In counting the number of possible arrangements of the five cards within a hand, there are $r = 5$ possible choices for the first position, then $r - 1 = 4$ possible choices for the second position, and so on. Thus, the number of possible arrangements within a five-card hand is $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5!$, and the total number of possible five-card arrangements that could be dealt from the 52-card deck is the product $(5!)(2,598,960) = 311,875,200$. Note that this is $(5!)C(52,5) = 52!/47! = P(52,5)$, the number of permutations that can be formed from 52 objects taken five at a time.

There is a natural relation between set theory and probability theory. In probability theory the universal set is the **sample space**, each subset of the sample space is an **event**, and an element is a **sample point**. With this understanding, probability is defined as follows:

Definition 2.10†

For any event A in a sample space S , the probability of occurrence of A , denoted by $\Pr(A)$, is a number associated with A such that

1. $0 \leq \Pr(A) \leq 1$
2. $\Pr(S) = 1$
3. If A and B are mutually exclusive events, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

There is no necessary relation between the theory of probability based on the axiomatic approach used here and the real world. With this approach, the theory of probability bears the same relation to the real world as, say, the theory of relativity. Both theories are logical constructs which purport to describe or predict the results of observations or experiments with physical objects. As with any other scientific theories, these theories are accepted as valid to the extent that they can adequately describe or predict the results of actual observations or experiments. The validity of the axiomatic approach is established by the fact that countless observations and experiments have demonstrated that real physical objects behave according to the descriptions and predictions of the theory of probability.

The analogy between the theory of probability and the theory of relativity can be taken further. The theory of relativity involves physical constants such as c , the speed of light. The theory does not specify the actual

†This definition of probability is not intended to be complete or mathematically rigorous. A complete exposition appears in Cramér (1946).

value of c , which must be determined by measurement. In the same way, the theory of probability does not provide the value of $\text{Pr}(A)$ —it must either be assumed or estimated. For example, suppose an experiment consists of throwing a die. Let $P(n)$ be the probability that the die shows n spots. The theory of probability only tells us that the $P(n)$ sum to one. In particular, it does *not* follow from the axioms that $P(n) = 1/6$ for every n . The numbers $P(n)$ should, in fact, be regarded as physical constants of the particular die that is being used. Their numerical values are not provided by the axioms of probability theory, any more than the size and the weight of the die are determined by the axioms of geometry and mechanics. Experience shows, however, that for a well-made die the frequency of any face coming up in a long series of throws usually approaches $1/6$ of the number of throws. Accordingly, it is often assumed that the die is “fair” and that all $P(n) = 1/6$. This is, however, an assumption and not a logical consequence of the axioms.

If the value of a probability is not assumed, then it must be estimated. The process of estimation of a probability is the exact analogue of the process of measurement used to obtain values for physical constants. Whether a value for $\text{Pr}(A)$ is assumed or estimated, the number obtained is only an approximation to the probability of the physical outcome corresponding to A . The true state of nature is never exactly characterized by theories or measurements.

The assumption that the faces of a die each have probability $1/6$ of coming up is a special case of an assumption that is often made. This assumption is that the sample points in a sample space are **equally likely**. If there are N equally likely sample points in the sample space, it follows from axioms (2) and (3) in Definition 2.10 that each has probability $1/N$.

Assume that the sample points in a sample space are equally likely and that A is any event. Then the **probability of occurrence of the event A** is the ratio of the number of sample points in the event A to the number of sample points in the sample space. Symbolically, this is $\text{Pr}(A) = n(A)/N$, where $n(A)$ is the number of sample points in the event A , and N is the number of sample points in the sample space. Also, $0 \leq \text{Pr}(A) \leq 1$.

In some treatments of probability, $n(A)/N$ is taken as the definition of $\text{Pr}(A)$. Although this is a possible approach when the sample points are assumed to be equally likely, this definition cannot be easily extended when the sample points are not equally likely. The axiomatic approach applies to all cases and yields $\text{Pr}(A) = n(A)/N$ as a theorem in the equally likely case.

Another common approach to probability is to identify the probability of an event A with its limiting relative frequency. If a large number M of experiments is performed and A occurs m times, then the observed relative frequency of A is m/M . This approach has a number of drawbacks. First, no matter how large M is, the observed relative frequency of A is, by

itself, a random quantity and is only an estimate of $\Pr(A)$. Second, it may not even be a good estimate, because even a fair coin can come up heads, say, 75 times in 100 tosses. Third, if the limiting value of m/M as $M \rightarrow \infty$ is used as the definition of $\Pr(A)$, the difficult conceptual problem remains of defining probability in terms of an experiment that is impossible to carry out in the real world.

The basis for the relative frequency approach to probability stems from the observation that, in countless experiments that have been performed, the observed relative frequency of an event A does seem to approach a limit. The counterpart to this observation is the law of large numbers, which is a theorem in the theory of probability.

Theorem 2.3 (Law of Large Numbers)

Let A be an arbitrary event in a sample space with probability $\Pr(A)$. If $m(A)$ is the number of occurrences of A in M experiments, then

$$\Pr(A) = \lim_{M \rightarrow \infty} \frac{m(A)}{M}$$

The following theorem is a direct analogy to Theorems 2.1 and 2.2.

Theorem 2.4

For arbitrary events A and B , $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$. If A and B are mutually exclusive, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Definition 2.11

Let A be an event in an arbitrary sample space with $\Pr(A) \neq 0$. Let B be any event in the same sample space. Then, the **conditional probability** that B occurs, when it is known that A has occurred, is defined by $\Pr(B|A) = \Pr(AB)/\Pr(A)$. Then by algebraic manipulation, $\Pr(AB) = \Pr(A)\Pr(B|A)$. Similarly, if $\Pr(B) \neq 0$, $\Pr(AB) = \Pr(B)\Pr(A|B)$.

Definition 2.12

Two events A and B are said to be **statistically independent** if $\Pr(A|B) = \Pr(A)$ or $\Pr(B|A) = \Pr(B)$.[†] From Definition 2.11, this is equivalent to saying that A and B are statistically independent if $\Pr(AB) = \Pr(A)\Pr(B)$.

To illustrate these concepts, consider the experiment of drawing one card from a shuffled deck of 52 playing cards. First, note that there are 52 cards (sample points) in the sample space and each card has an equal chance of being the one selected. Define A to be the event that the

[†] From Definition 2.11, if either of these equations holds, then the other must also hold.

selected card is a heart and B to be the event that the selected card is a king. From Definition 2.10 and the previous discussions, $\Pr(A) = 13/52 = 1/4$, $\Pr(B) = 4/52 = 1/13$, and $\Pr(AB) = 1/52$. Using Theorem 2.3, the probability of selecting a heart or a king or both is $\Pr(A \cup B) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13$.

Now suppose that in this experiment, a card is selected by someone who tells us that it is a heart but does not tell us the face value. What is the probability that this card is a king? Using Definition 2.11, the probability it is a king, given that it is a heart, is $\Pr(B|A) = \Pr(AB)/\Pr(A) = (1/52)/(13/52) = 1/13$.

Notice that in this example $\Pr(B|A) = \Pr(B) = 1/13$; i.e., knowing that the selected card is a heart does not alter the probability that it is a king. Thus, by Definition 2.12, the events A and B are statistically independent. The practical implication is that knowing a card is a heart provides no additional information about its possible face value.

Suppose now that a "joker" is added to the deck (sample space). The experiment and the events A and B are as defined above but now $\Pr(A) = 13/53$, $\Pr(B) = 4/53$, and $\Pr(AB) = 1/53$. The probability of a king, given that a heart was selected, is $\Pr(B|A) = (1/53)/(13/53) = 1/13$ as before; however, $\Pr(B) \neq \Pr(B|A)$ (i.e., $4/53 \neq 1/13$). Also note that $\Pr(A) \cdot \Pr(B) \neq \Pr(AB)$ [i.e., $52/(53)^2 \neq 1/53$]. Thus, events A and B are not statistically independent when a joker is added to the deck. In this case, knowing that a card is a heart does provide additional information about its possible face value (i.e., we know it is not the joker).

2.2 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

In the previous section the concept of an outcome of an experiment is introduced. In most applications it is desirable to assign numbers to the possible outcomes. Usually this occurs quite naturally. For example, outcomes can be the percent ^{235}U measured in a sample, the number of gross discrepancies between the facility and an audit team in an inspection situation, or the amount of plutonium in a barrel of solid waste. When a number is not assigned naturally to an outcome, one can be assigned to represent it. Thus, for example, if we speak of an attempted diversion, either there will be an attempt of some kind in a given instance or there will not be. For convenience, the values 1 and 0 can be arbitrarily assigned to the two possible outcomes. This assignment of numbers is described by defining a random variable.

A **random variable** is a numerically valued function defined over a sample space. It is a rule which assigns a numerical value to each possible outcome of an experiment.

As an example, consider the simple experiment of rolling two dice. Define an outcome to be the numbers facing up on the two dice after they are rolled. Let (a,b) denote observing the numbers a on the first die and b on the second. Table 2.1a lists the sample space for this experiment, which consists of 36 possible outcomes (sample points).

Suppose the sum of the numbers observed on the two dice is of interest. The random variable $X = a + b$ is defined, which assigns a number to each sample point. These are listed in Table 2.1b.

TABLE 2.1a

**The Sample Space for the Experiment
of Rolling Two Dice**

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

TABLE 2.1b

**The Random Variable X Assigns the Number
 $a + b$ to Each Sample Point**

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Perhaps the actual value of the sum $a + b$ is not important but simply whether or not $a + b$ is greater than 9. In this case, the random variable Y is defined such that $Y = 0$ if $a + b \leq 9$ and $Y = 1$ if $a + b > 9$. A random variable of this type is called an **indicator variable**. Y can be interpreted as follows. For each sample point, count the number of values of $a + b$ that are greater than 9. Because there is only one value of $a + b$ for each sample point, this count is either 0 or 1. Table 2.1c shows the values of Y assigned to the sample points.

TABLE 2.1c

The Random Variable Y Assigns 0
if $a + b \leq 9$ and 1 if $a + b > 9$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	1
0	0	0	0	1	1
0	0	0	1	1	1

The presentation and examples thus far may have left the impression that it is necessary to be able to completely list the sample space and to assign only integer values to the sample points based on some characteristic of interest. In practice, however, the sample points might be, for example, containers of UO_2 powder, and the characteristic of interest might be the weight of each container or the percent uranium of the powder in each container. Clearly, a random variable which assigns an integer value to a weight or a measurement is not wholly satisfactory for most applications. A more realistic random variable would take on a continuum of values in a specified interval.

There are two basic types of random variables, **discrete** and **continuous**. A **discrete random variable** is one that can take on only a finite or a denumerable number of values which are usually, but not necessarily, integers. In this context denumerable means that the possible values of a random variable can be put in one-to-one correspondence with the integers (counting numbers). A **continuous random variable** is one that can take on a continuum of values.

In most experiments, interest is centered on the values of a random variable rather than on all possible outcomes. As illustrated in Tables 2.1b and 2.1c, a single value of a random variable may be associated with more than one of the sample points. Thus, the set of sample points associated with a single value of a random variable is an event. The notation $X = x$ is used to denote the event that the random variable X takes on the value x .

Of course, it is rarely known in advance exactly what the outcome of a single trial of an experiment will be. (An exception is when, say, a two-headed coin is tossed.) It is often possible, however, to make assumptions about the experiment which allow the assignment of probabilities to the values of the random variable of interest. For the experiment of rolling two dice, the assumptions made are that each die has six sides which are numbered 1 through 6, the dice are fair (i.e., not weighted or rolled so as to favor any particular outcome), and that the outcomes are statistically independent. Under these assumptions, a probability of $1/36$ is assigned to

each of the individual sample points listed in Table 2.1a. Definition 2.10 is used to assign probabilities to the possible values of the random variable X in Table 2.1b. These are listed in Table 2.2. Similarly, if probabilities are assigned to the values of the random variable Y in Table 2.1c, then $\Pr(Y = 0) = 5/6$ and $\Pr(Y = 1) = 1/6$.

An enumeration of the possible values of a random variable with their associated probabilities, such as Table 2.2, is called the **probability distribution** of the random variable. For the example, note from Table 2.2 that on a single trial of rolling two dice, $X = 7$ is more likely to be observed than any other single value. Because the events $X = x$ are mutually exclusive for $x = 2, 3, \dots, 12$, Theorem 2.3 can be implemented to compute, for example, $\Pr(X \leq 3) = \Pr(X = 2) + \Pr(X = 3) = 1/36 + 2/36 = 3/36$.

TABLE 2.2
Pr($X = x$) for the Experiment
of Rolling Two Dice*

x	Pr($X = x$)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

* X is the total of the two face values.

In the previous example, the probabilities were obtained from a complete listing of the sample space. Such a listing, however, is not always possible in practice, and even if it were possible, it usually is not practical. A more realistic approach is to define a **probability model** which expresses $\Pr(X = x)$ as a function of x over the range of possible values of x . Although this concept is not directly applicable to continuous random variables, the analogy should become clear as the presentation progresses.

Conventional concepts and notations are now given for assigning probabilities to the possible values of a random variable.

For any random variable X ,

$$F(x) = \Pr(X \leq x) \quad (2.1)$$

is called the **cumulative distribution function (cdf)** of X . If X is a **discrete** random variable, the **probability function (pf)** of X is defined to be

$$f(x) = \Pr(X = x) \quad (2.2)$$

and the cdf is given by

$$F(x) = \sum_{y \leq x} f(y) \quad (2.3)$$

If X is a **continuous** random variable with a differentiable cdf, the **probability density function (pdf)** is defined to be

$$f(x) = dF(x)/dx \quad (2.4)$$

and the cdf is given by

$$F(x) = \int_{-\infty}^x f(y) dy \quad (2.5)$$

It is common practice to distinguish between pf's and pdf's as corresponding to discrete and continuous random variables, respectively. For convenience of presentation, however, pdf is used to refer to either a pf or a pdf throughout the remainder of this book.

Theorem 2.5

A pdf $f(x)$ has the following properties:

1. $f(x) \geq 0$
2. $\sum_{\text{all } x} f(x) = 1$ if X is discrete
3. $\int_{-\infty}^{\infty} f(x) dx = 1$ if X is continuous.

Theorem 2.6

If X is discrete,

$$\Pr(a < X \leq b) = F(b) - F(a) = \sum_{a < x \leq b} f(x)$$

If X is continuous,

$$\Pr(a < X < b) = \Pr(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$$

It should be noted from Theorem 2.6 that if X is continuous, then strictly speaking, the probability that X is exactly equal to x is 0. Because of limitations on our ability to measure and record x in practice, however, we usually talk about an interval ($x - \Delta < X < x + \Delta$). For example, the weights of UO_2 powder containers may be rounded to the nearest 10 g. Thus, a weight recorded as 22.03 kg could actually be interpreted as being from the interval ($22.025 \text{ kg} < X < 22.035 \text{ kg}$). The probability of observing a value of X from this interval is $F(22.035) - F(22.025)$, which is not zero unless $F(22.035) = F(22.025)$.

In Sections 2.4 and 2.5 some specific pdf's are presented that have applications in nuclear material accounting and are the basis for the statistical methods presented throughout this book.

2.3 EXPECTED VALUES

Although the probability distribution of a random variable is completely specified by its pdf, it is often convenient to work with some descriptive characteristics of the random variable. This section introduces descriptive measures which are expected values of certain functions of the random variable.

The **expected value** of any function of a random variable is defined as the weighted average (weighted by the probability of its occurrence) of the function over all possible values of the random variable. The symbol $E[\cdot]$ is used to denote the expected value of whatever appears within the brackets. For example, the expected value of a function $u(X)$ is denoted by $E[u(X)]$. To simplify the notation in subsequent discussions, the expression \sum_x denotes summation over all values of x .

For a random variable X , the expected value of a function $u(X)$ is defined as

$$E[u(X)] = \sum_x u(x) f(x) \quad (2.6)$$

if X is discrete, or

$$E[u(X)] = \int_{-\infty}^{\infty} u(x) f(x) dx \quad (2.7)$$

if X is continuous.

The **mean** and **variance** of a random variable X are special cases of $E[u(X)]$. The **mean** provides a measure of central tendency in X and is defined as

$$\mu_x = E[X] = \sum_x xf(x) \quad (2.8)$$

if X is discrete, or

$$\mu_x = E[X] = \int_{-\infty}^{\infty} xf(x) dx \quad (2.9)$$

if X is continuous. When it is clear from the context that the random variable is X , μ_x is denoted by μ .

The **variance** describes the spread or dispersion of the possible values of X about the mean and is defined as $E[(X - \mu)^2]$; i.e.,

$$\sigma_x^2 = \text{Var}[X] = \sum_x (x - \mu)^2 f(x) \quad (2.10)$$

if X is discrete, or

$$\sigma_x^2 = \text{Var}[X] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad (2.11)$$

if X is continuous. When it is clear from the context that the random variable is X , σ_x^2 is denoted by σ^2 . An alternative form of the variance which is sometimes more convenient to evaluate is derived as

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 \end{aligned} \quad (2.12)$$

for either a discrete or continuous random variable. This derivation of Equation 2.12 involves linearity properties of expected values that are not presented here.

The variance is the average squared deviation from the mean. Thus, for example, if the units of measure for the random variable X are grams,

then the units for the variance would be grams squared. Another measure of dispersion, the **standard deviation**, is defined as the positive square root of the variance and is given by

$$\sigma = \sqrt{\sigma^2} \quad (2.13)$$

The standard deviation is in the same units of measure as X and is often more easily interpreted than the variance.

The mean and standard deviation are sometimes referred to as **location** and **scale** parameters. It is a common practice in statistics to express a random variable as a distance from its mean in units of its standard deviation. This is accomplished by the transformation

$$Z = \frac{X - \mu}{\sigma} \quad (2.14)$$

This transformation **standardizes** X . The random variable Z has a mean equal to zero and a variance equal to one. This concept is used and further discussed in subsequent sections and chapters.

In the previous discussion the terminology “mean and variance of a random variable X ” was used. Another commonly used terminology is “mean and variance of a distribution.” Either terminology is acceptable, and they are used interchangeably throughout this text.

An expectation of the form $\mu_k = E[X^k]$ is called the **k^{th} moment** of the distribution of X . For example, the mean is the first moment $\mu_1 = E[X]$, and the variance, given by Equation 2.12, is $\sigma^2 = \mu_2 - \mu_1^2$ or the second moment minus the square of the first moment.

An expectation of the form $E[(X - \mu)^k]$ is known as the **k^{th} moment about the mean** and is denoted by μ'_k . For example, the variance is μ'_2 , the second moment about the mean. The third moment about the mean is $\mu'_3 = E[(X - \mu)^3]$, which is a measure of **skewness**, or lack of symmetry about the mean, of a pdf. The fourth moment about the mean is $\mu'_4 = E[(X - \mu)^4]$, which is a measure of the **kurtosis**, or peakedness, of the pdf. The quantity

$$\gamma_1 = \frac{\mu'_3}{\sigma^3} \quad (2.15)$$

is a standardized measure of skewness, and

$$\gamma_2 = \frac{\mu'_4}{\sigma^4} \quad (2.16)$$

is a standardized measure of kurtosis.

2.4 CERTAIN DISCRETE DISTRIBUTIONS

In this section, three discrete distributions are presented that are widely used in nuclear material accounting. They are the hypergeometric, binomial, and Poisson distributions. These distributions are commonly used to compute probabilities for attribute sampling plans (introduced in Chapter 17).

2.4.1 Hypergeometric Distribution

Suppose that a collection of N objects contains D of one type and $N - D$ of another type and that exactly $n \leq N$ objects are randomly selected without replacement from the collection. Let the random variable X be the number of objects of the first type that appear in the sample. The probability that $X = x$ is computed from the **hypergeometric pdf**, which is given by

$$f(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad (2.17)$$

$$x = 0, 1, \dots, \min(D, n)$$

where $\min(D, n)$ indicates the minimum or smaller of the values D and n , and where

$$\binom{a}{b} = C(a, b) = \frac{a!}{b!(a-b)!}$$

is the number of possible combinations of a items taken b at a time (Definition 2.9).

The mean and variance of the hypergeometric distribution are given, respectively, by

$$\mu = E[X] = \frac{nD}{N} \quad (2.18)$$

and

$$\sigma^2 = \frac{(N-n)}{(N-1)} \cdot \frac{nD(N-D)}{N^2} \quad (2.19)$$

In nuclear material accounting, the hypergeometric distribution is applicable to certain audit inspection activities. In this context, N might be the number of containers of a particular type that are on inventory. Then D would be the number of these containers that possess some characteristic of interest. The D items are typically called “defectives” or “discrepant,” because the characteristic of interest is often a difference of a given size between the operator’s statement of what is in a container and what an audit team finds. Then n of the containers are randomly selected without replacement from the N containers and examined by the audit team. If X is the number of defectives observed in the sample, then the probability that $X = x$ is given by Equation 2.17. The mean, given by Equation 2.18, is interpreted as the number of defectives we would expect to observe in a sample of size n when D of the N containers are defectives.

Example 2.1 Assume that $N = 50$ items are on inventory and that $n = 25$ of these items are randomly selected without replacement by an audit team. Each selected item is classified as either being defective (in some sense) or not being defective. The inventory is considered acceptable by the audit team if $x \leq 1$ defective items are observed in the sample. What is the probability of accepting the inventory when the total number of defective items on inventory is $D = 2$, $D = 5$, or $D = 10$?

The solution uses Equations 2.1, 2.3, and 2.17. The probability of acceptance is

$$\begin{aligned} \Pr(X \leq 1) &= \Pr(X = 0 \text{ or } X = 1) \\ &= \Pr(X = 0) + \Pr(X = 1) = f(0) + f(1) \\ &= \frac{\binom{D}{0} \binom{50-D}{25-0}}{\binom{50}{25}} + \frac{\binom{D}{1} \binom{50-D}{25-1}}{\binom{50}{25}} \end{aligned}$$

For $D = 2$, the probability of acceptance is

$$\begin{aligned} \Pr(X \leq 1) &= \frac{\binom{2}{0} \binom{48}{25}}{\binom{50}{25}} + \frac{\binom{2}{1} \binom{48}{24}}{\binom{50}{25}} \\ &= 0.24490 + 0.51020 = 0.75510 \end{aligned}$$

For $D = 5$,

$$\begin{aligned}\Pr(X \leq 1) &= \frac{\binom{5}{0}\binom{45}{25}}{\binom{50}{25}} + \frac{\binom{5}{1}\binom{45}{24}}{\binom{50}{25}} \\ &= 0.02508 + 0.14926 = 0.17434\end{aligned}$$

For $D = 10$,

$$\begin{aligned}\Pr(X \leq 1) &= \frac{\binom{10}{0}\binom{40}{25}}{\binom{50}{25}} + \frac{\binom{10}{1}\binom{40}{24}}{\binom{50}{25}} \\ &= 0.000318 + 0.004972 = 0.00529\end{aligned}$$

Thus, if 2 of the 50 items on inventory are defectives, there is a 76% chance (i.e., the probability is 0.76) that the inventory will be accepted. If 5 of the 50 items are defectives, there is a 17% chance that the inventory will be accepted. If 10 of the 50 items are defectives, there is a 0.5% chance that the inventory will be accepted.

Suppose that the manager of the audit team feels that when $D = 5$, a 17% chance of accepting the inventory is too large. The probabilities of accepting the inventory for a given value of D can be decreased by increasing the sample size n . For example, if $n = 30$ and $D = 5$, $\Pr(X \leq 1) = 0.07592$. That is, the probability of accepting the inventory when $D = 5$ can be decreased from 0.17434 to 0.07592 by increasing the sample size from 25 to 30.

The number of defectives we would expect to observe when $N = 50$, $n = 25$, and $D = 5$ is computed from Equation 2.18 to be

$$\mu = \frac{25(5)}{50} = 2.5$$

The variance is computed from Equation 2.19 to be

$$\sigma^2 = \frac{(50-25)}{(50-1)} \cdot \frac{25(5)(50-5)}{50^2} = 1.148$$

and the standard deviation is $\sigma = \sqrt{1.148} = 1.07$.

2.4.2 Binomial Distribution

Suppose that an experiment can have only two possible outcomes, conveniently called “success” and “failure,” and suppose that the probability of a success is p . If the experiment is repeated independently n times under exactly the same conditions and if the random variable X is defined to be the number of successes observed in the n **independent trials**, then X has a **binomial distribution**. Before formally presenting the binomial pdf, it is instructive to consider an application.

In the inventory example of the previous section, suppose that the sample is selected **with replacement**. Theoretically, when sampling with replacement, an item is randomly selected from the N items on inventory, is classified as either being defective or not being defective, and is returned to the inventory. This process of random selection, classification, and replacement is repeated n times. It is possible for any item to be in the sample more than once. Of course when this happens in practice, the item is not actually measured repeatedly. Instead, the item is measured once, and this measured value is reported each time the item is in the sample. Under this sampling plan, the n classifications are equivalent to n **independent trials** of an experiment. This is because the probability of a defective occurring ($p = D/N$) is the same for every observation, regardless of the outcomes of previous observations (this is *not* true for sampling without replacement).

The **binomial pdf** is given by

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n \quad (2.20)$$

The mean of X is given by

$$\mu = E[X] = np \quad (2.21)$$

and the variance of X is given by

$$\sigma^2 = E[(X - \mu)^2] = np(1 - p) \quad (2.22)$$

Although $f(x)$, given by Equation 2.20, can be programmed on a computer or hand-held calculator, Table A1 of Appendix A provides values of $F(x) = \Pr(X \leq x)$ for n ranging from 1 to 20 and p ranging from 0.05

to 0.95 in increments of 0.05 (i.e., 0.05, 0.10, 0.15, ..., 0.95). For example, if $n = 10$ and $p = 0.30$, suppose that $\Pr(X \leq 4)$ is to be determined. Enter Table A1 with $n = 10$ and $x = 4$ in the left margin. Follow this row of the table to the column headed by $p = 0.30$. The table gives $\Pr(X \leq 4) = 0.8497$.

Table A1 can also be used with Theorem 2.6 to determine individual probabilities of the form $f(x) = \Pr(X = x)$ by using the relationship $\Pr(X = x) = \Pr(X \leq x) - \Pr(X \leq x - 1)$. In the above example, $\Pr(X = 4) = \Pr(X \leq 4) - \Pr(X \leq 3) = 0.8497 - 0.6496 = 0.2001$.

Example 2.2 Assume that sampling with replacement is implemented in Example 2.1. The sample size is $n = 25$, and the probability of a success (defective) is $p = D/N = D/50$ for each of the 25 observations. The inventory will be accepted if $x \leq 1$ defectives are observed. Thus, the probability of accepting the inventory is given by

$$\begin{aligned} \Pr(X \leq 1) &= \Pr(X = 0) + \Pr(X = 1) \\ &= \binom{25}{0} \left(\frac{D}{50}\right)^0 \left(1 - \frac{D}{50}\right)^{25} + \binom{25}{1} \left(\frac{D}{50}\right)^1 \left(1 - \frac{D}{50}\right)^{24} \end{aligned}$$

For $D = 2$, the probability of acceptance is

$$\Pr(X \leq 1) = (0.96)^{25} + 25(0.04)(0.96)^{24} = 0.73581$$

For $D = 5$,

$$\Pr(X \leq 1) = (0.9)^{25} + 25(0.1)(0.9)^{24} = 0.27121$$

For $D = 10$,

$$\Pr(X \leq 1) = (0.8)^{25} + 25(0.2)(0.8)^{24} = 0.02739$$

These probabilities of accepting the inventory are noticeably different, especially for $D = 5$ and $D = 10$, from the corresponding probabilities

computed in Example 2.1, where sampling without replacement was assumed.

It was pointed out previously that when sampling with replacement, any item could appear in the sample more than once. Thus, for a given sample size n , the sample collected with replacement would be expected to have less information about the inventory (i.e., fewer distinct items) than the sample collected without replacement. For this reason, sampling with replacement is less efficient than sampling without replacement for some applications.

When N is "large" relative to n , however, the binomial pdf provides a satisfactory approximation to the hypergeometric pdf for computing probabilities when sampling is without replacement. A widely accepted definition of "large" is $N > 50$ and $n/N \leq 0.10$, where the approximation gets better as N approaches ∞ and n/N approaches zero. Rohatgi (1984) provides a mathematically rigorous treatment of this subject. In such cases, sampling without replacement does not significantly alter the sample space from one trial to the next. Thus, the probability of a success stays practically constant for all trials, and the conditions for using the binomial distribution are reasonably well satisfied.

To illustrate this, consider an example of sampling without replacement where $N = 1000$, $D = 200$, and $n = 10$. In this case $n/N = 10/1000 = 0.01$. For the first item selected, the probability of observing a defective is $p = 200/1000 = 0.2$. The value of p for the second item selected, however, depends on the outcome of the first trial. If the first item selected is a defective, then 199 of the remaining 999 items are defectives, and the probability of observing a defective on the second trial is $p = 199/999 = 0.1992$. If the first item selected is not a defective, then 200 of the remaining 999 items are defectives and $p = 200/999 = 0.2002$ for the second trial. In either case, the outcome of the first trial has only slightly changed the value of p for the second trial.

Now suppose that the first nine items selected are all defectives. Then the probability of a defective on the 10th trial is $p = 191/991 = 0.1927$. On the other hand, if the first nine items selected are all nondefectives, then $p = 200/991 = 0.2018$ for the 10th trial. Because these are the two extremes of the possible outcomes, the value of p is within the interval $0.1927 \leq p \leq 0.2018$ for every trial. Thus, for practical purposes, the assumption of independent trials (i.e., constant p for all trials) is not too badly violated, and the binomial pdf with $p = 0.2$ and $n = 10$ provides an acceptable approximation to the hypergeometric pdf.

This concept is further illustrated by comparing the means and variances of the hypergeometric and binomial distributions. By letting $p =$

D/N , the mean of the hypergeometric distribution given by Equation 2.18 can be expressed as

$$\mu = nD/N = np \quad (2.23)$$

which is the mean of the binomial distribution given by Equation 2.12. Thus, the two distributions have the same mean. Substituting $p = D/N$ into Equation 2.19, the variance of the hypergeometric distribution can be expressed as

$$\sigma^2 = \left(\frac{N-n}{N-1} \right) np(1-p) \quad (2.24)$$

which is the variance of the binomial distribution multiplied by $(N-n)/(N-1)$.

The quantity $(N-n)/(N-1)$ is called the **finite population correction factor**. As N approaches ∞ (for $n > 1$), the variances of the two distributions are the same. When $n/N = 0.10$, the finite population correction factor is slightly less than 0.90, so that the variance of X is inflated by about 11% and the standard deviation by about 5% when the binomial distribution is used to approximate the hypergeometric distribution. For most practical applications, this amount of inflation is not considered to be serious and may often be ignored.

2.4.3 Poisson Distribution

The **Poisson distribution** is applicable for computing probabilities of events in time and space. Its development is best illustrated by an example. Consider a nondestructive assay. The item being assayed will emit gamma particles at a rate (say λ per unit of time) that depends on the amount of one or more nuclides in the item. Suppose that the emitted particles are counted over a finite time interval T . Let the time interval T be divided into a large number $n = T/\tau$ of arbitrarily small intervals of length τ . In one of these small intervals, the probability that a particle reaches the counter is $p = \lambda\tau$, irrespective of whether previous particles have been counted recently. That is, p is assumed to be constant from one small interval to another. Thus, the number of particles counted in time T has a binomial distribution with $n = T/\tau$ and $p = \lambda\tau$. Now if in Equation 2.20 we let $n \rightarrow \infty$ and $p \rightarrow 0$ while holding np constant, the result is the **Poisson pdf**, which is given by

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, \dots, \quad (2.25)$$

where μ is the constant value of np . For the Poisson distribution, both the mean and the variance are equal to μ . That is,

$$E[X] = E[(X - \mu)^2] = \mu \quad (2.26)$$

It may seem unusual that the mean and variance are equal. The variance of the binomial distribution is $np(1 - p)$, however, and if p approaches 0 while np remains constant, then $(1 - p)$ approaches 1, and $np(1 - p)$ approaches $np = \mu$.

In the nondestructive assay example, as $\tau \rightarrow 0$, note that $n = T/\tau \rightarrow \infty$ and $p = \lambda\tau \rightarrow 0$, while $\mu = np = \lambda T$ remains constant. Thus, as τ becomes very small, the number of particles counted in time T has a Poisson distribution with pdf given by Equation 2.25 where $\mu = \lambda T$.

Similarly, if particles of a certain type are distributed randomly in a liquid with density λ per unit volume, the number of particles found in a sample of volume V has a Poisson distribution with $\mu = \lambda V$.

Although the Poisson pdf, $f(x)$, given by Equation 2.25, is easily programmed on a computer or hand-held calculator, Table A2 of Appendix A provides values of $\Pr(X \leq x)$ for selected values of μ ranging from 0.01 to 20. For example, suppose that $\mu = 3.0$, and the value of $\Pr(X \leq 4)$ is to be determined. Enter Table A2 and find the column headed by $\mu = 3.0$. Follow this column down to the row with $x = 4$ in the left margin. The table gives $\Pr(X \leq 4) = 0.815$.

The value of $\Pr(X = x)$ is determined from Table A2 by taking the difference $\Pr(X \leq x) - \Pr(X \leq x - 1)$. For the above example, $\Pr(X = 4) = \Pr(X \leq 4) - \Pr(X \leq 3) = 0.815 - 0.647 = 0.168$.

The development of the Poisson pdf as a limiting form of the binomial implies that, under certain conditions, the Poisson pdf could be used to approximate the binomial. There are a number of proposed "rules of thumb" that indicate when the Poisson provides an acceptable approximation to the binomial. Many texts state that if $n \geq 20$ and $p \leq 0.05$, the approximation is acceptable, and that it is very good if $n \geq 100$ and $np \leq 10$.

In some audit inspection applications of attribute sampling, the number of defective items on inventory considered tolerable is quite small relative to the total number of items on inventory. In such cases, it is necessary to inspect a large sample of the items to verify the inventory. For such situations, it is sometimes possible to approximate the hypergeometric distribution with the Poisson. This requires that $n/N \leq 0.10$, so that the binomial can be used to approximate the hypergeometric, and it also requires that

the rules of thumb for approximating the binomial with the Poisson be satisfied.

Example 2.3 Consider an audit inspection situation where D of the $N = 250$ items on inventory are defective. A random sample of $n = 25$ items will be selected without replacement. The inventory will be accepted if $x \leq 1$ defectives are observed in the sample. Suppose that the probability of accepting the inventory when $D = 10$ is to be evaluated. Because $n/N = 0.10$, the binomial distribution with $p = D/N = 0.04$ and $n = 25$ can be used to approximate the hypergeometric. Also, because $p = 0.04 < 0.05$ and $n = 25 > 20$, the Poisson distribution with $\mu = np = 1$ can be used to approximate the binomial (and hence the hypergeometric).

For comparison, Table 2.3 displays the probability distributions computed from Equation 2.25 with $\mu = 1$, Equation 2.20 with $p = 0.04$ and $n = 25$, and Equation 2.17 with $N = 250$, $D = 10$, and $n = 25$. For this example, the Poisson approximation to the binomial is quite accurate,

TABLE 2.3
Probabilities Computed from
Poisson, Binomial, and
Hypergeometric Distributions

x	Poisson $\mu = 1$	Binomial $n = 25,$ $p = 0.04$	Hypergeometric $N = 250,$ $D = 10, n = 25$
0	0.3679	0.3604	0.3416
1	0.3679	0.3754	0.3954
2	0.1839	0.1877	0.1968
3	0.0613	0.0600	0.0554
4	0.0153	0.0137	0.0097
≥ 5	0.0037	0.0028	0.0011

and for practical purposes the Poisson approximation to the hypergeometric is probably acceptable. The Poisson approximation of the probability of accepting the inventory is given by

$$\Pr(X \leq 1) = 0.3679 + 0.3679 = 0.7358$$

This is quite close to the exact probability of acceptance, which is computed from the hypergeometric distribution to be

$$\Pr(X \leq 1) = 0.3416 + 0.3954 = 0.7370$$

For most practical situations, this would be rounded to two significant digits and presented as 0.74, in which case the Poisson approximation is identical to the exact value.

2.5 CERTAIN CONTINUOUS DISTRIBUTIONS

In this section some selected continuous distributions are presented. The uniform, normal, and lognormal distributions are commonly used in nuclear material accounting applications and are discussed in some detail. Other distributions, which have potential applications in nuclear material accounting, are briefly introduced.

2.5.1 Uniform Distribution

The **uniform distribution** has a variety of applications in nuclear material accounting. These include modeling the rounding errors in weighing processes, randomization in designed experiments and sampling plans (Chapters 7 and 8), and Monte Carlo simulation techniques (Chapter 11).

The uniform probability density function is given by

$$\begin{aligned} f(x) &= \frac{1}{\beta - \alpha}, \quad \alpha \leq x \leq \beta \\ &= 0, \quad \text{otherwise,} \end{aligned} \tag{2.27}$$

where α and β are real constants with $\alpha < \beta$. The uniform pdf is illustrated in Figure 2.2. Because a uniformly distributed random variable has a pdf that is constant over the interval $\alpha \leq x \leq \beta$, the pdf is simply the reciprocal of the length of the interval.

The mean and variance of the uniform distribution are given by

$$\mu = E[X] = \frac{\alpha + \beta}{2} \quad (2.28)$$

and

$$\sigma^2 = E[(X - \mu)^2] = \frac{(\beta - \alpha)^2}{12} \quad (2.29)$$

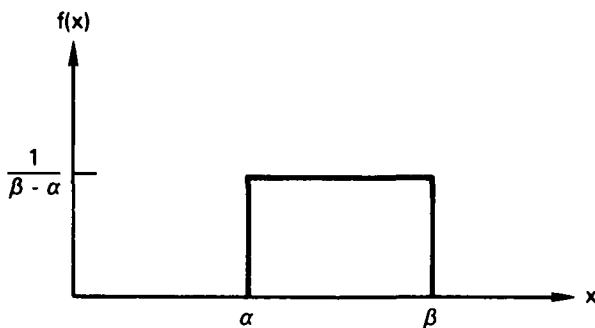


Figure 2.2 A uniform pdf with parameters $\alpha < \beta$.

The cdf, $F(x) = \Pr(X \leq x)$, is given by

$$\begin{aligned} F(x) &= 0, & x < \alpha \\ &= \frac{x - \alpha}{\beta - \alpha}, & \alpha \leq x \leq \beta \\ &= 1, & x > \beta \end{aligned} \quad (2.30)$$

For any subinterval $a \leq X \leq b$, where $\alpha \leq a \leq b \leq \beta$, the probability $\Pr(a \leq X \leq b)$ is given by

$$\Pr(a \leq X \leq b) = F(b) - F(a) = \frac{b - a}{\beta - \alpha} \quad (2.31)$$

and this probability is the same for all such subintervals of length $b - a$.

Example 2.4 Consider the effect of rounding off measurement data. Suppose that numbers of the form NN.N are “rounded off” to the nearest integer. One round-off procedure is such that if the decimal part is less

than 0.5, the number is rounded down by simply dropping the decimal part. If the decimal part is 0.5 or larger, the number is rounded up by adding 1 and then dropping the decimal part. The round-off error X is defined as the difference between the number before rounding and the number after rounding. These errors are commonly assumed to have a uniform distribution on the interval $[-0.5, 0.5]$. That is,

$$f(x) = \frac{1}{0.5 - (-0.5)} = 1 \quad \text{when } -0.5 \leq x \leq 0.5$$

$$= 0, \quad \text{otherwise.}$$

The mean and variance of X are given by

$$\mu = (-0.5 + 0.5)/2 = 0.0$$

and

$$\sigma^2 = [0.5 - (-0.5)]^2/12 = 1/12$$

Suppose that the numbers being rounded are observed values of a random variable, say Y . The rounded numbers are then observed values of the sum of two random variables, say $Z = Y + X$. If Y and X are independent, the effect of rounding to the nearest integer is that the variance of Z is larger than the variance of Y . The difference is the variance of X , which is $1/12$. (The variance of a sum of random variables is discussed in Section 2.7.1.)

Jaech (1973) states that round-off error may be a dominant source of the variation due to weighing some types of items. He discusses the effect of rounding recorded weights to the nearest 50 g and then gives the variance and standard deviation of the round-off error for six rounding rules. These results are summarized in Table 2.4.

Probably the most common use of the uniform distribution is in applications of Monte Carlo simulation modeling, as discussed in Chapter 11. Briefly, all computers and many hand-held calculators can be programmed to generate the equivalent of a random sample from a uniform distribution with specified parameters. These values can then be transformed to simulate a random sample from a specified discrete or continuous distribution.

Thus, computerized models can be developed to simulate complicated processes. This allows study of the effects on a process due to modifying various operating conditions. It is then feasible to conduct studies that would otherwise be extremely difficult.

TABLE 2.4
Variance Due to Rounding

Round to nearest K units*	Variance in (units) ²	Standard deviation
K = 1	0.0833	0.29
5	2.0833	1.44
10	8.3333	2.89
25	52.0833	7.22
50	208.3333	14.43
100	833.3333	28.87

*Units may be g, kg, lb, or other units of weight.

In the simulation context, the uniform distribution is used to generate randomization schemes for sampling plans (Chapter 8) and for experimental designs (Chapter 7). As a simple example, suppose that a list is available of 1000 items on inventory and that 25 of these items are to be selected for inspection. Assume that sampling with replacement is desired. The solution is to generate a random sample of size 25 from the uniform distribution with parameters $\alpha = 0.50$ and $\beta = 1000.49$. The resulting 25 values are then rounded to the nearest whole integer and arranged in ascending order. Note that $\alpha = 0.50$ ensures that, when rounding to the nearest whole integer, the value 1 is the smallest possible. Similarly, $\beta = 1000.49$ ensures that integers larger than 1000 are never generated. This provides a list of the randomly selected items to be inspected. Specifically, suppose the first two integers on the sample list were 55 and 97. Then the 55th and 97th items on the inventory list would be two of the items scheduled for inspection. Of course, if any item appears in the sample more than once, it is inspected only once, and the inspection result would be recorded repeatedly (or simply weighted) according to the number of times the item is in the sample.

2.5.2 Normal Distribution

The **normal distribution** is undoubtedly the most commonly used distribution in statistical applications. A random variable X has a normal distri-

bution with mean μ and variance σ^2 if it has the probability density function given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty \quad (2.32)$$

The normal distribution is used so extensively that the shorthand notation $X \sim N(\mu, \sigma^2)$ has been adopted to indicate that the random variable X is normally distributed with mean μ and variance σ^2 .

The normal pdf, given by Equation 2.32, defines a “bell-shaped” curve which is symmetric about the mean μ . This is graphically illustrated in Figure 2.3.

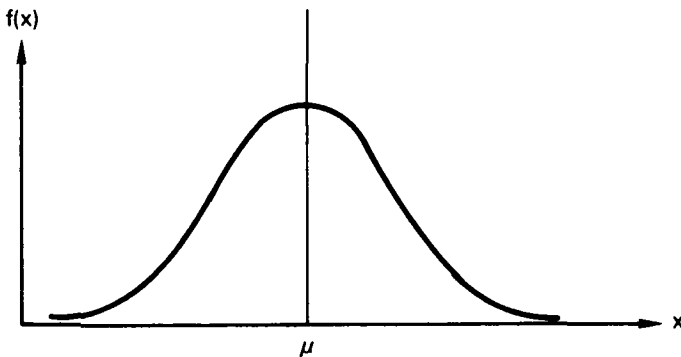


Figure 2.3 The normal pdf.

The normal distribution is used extensively in practice because many (but of course not all) random variables are either approximately normally distributed or can be made nearly so by a simple transformation. It is also the limiting distribution for many sample statistics. That is, as the sample size becomes large, many statistics computed from sample data tend to be normally distributed. This is discussed further in Chapters 3 and 4. Thus, many important statistical inference procedures are developed by assuming that the random variable of interest is normally distributed. Fortunately, most of these so-called “normal theory” procedures are not very sensitive to even moderate departures from normality. This means that, in practice, the assumption that a random variable is normally distributed can be violated to some moderate degree without seriously affecting the validity of conclusions reached when these procedures are applied to sample data.

The cumulative distribution function for the normal is given by

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt \quad (2.33)$$

This integral cannot be evaluated directly without resorting to numerical methods. Also, tabulating $F(x)$ for every possible pair of values (μ, σ) would be an impossible task. The random variable X can be standardized, however, by the transformation

$$Z = \frac{X - \mu}{\sigma} \quad (2.34)$$

This transforms X to the random variable Z which has a normal distribution with mean 0 and variance 1. This transformation brings all normal distributions to the same form, $N(0,1)$, which is called the **standard normal distribution**. $F(x)$ can then be evaluated by noting that

$$\begin{aligned} F(x) &= \Pr(X \leq x) = \Pr\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\ &= \Pr(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt \end{aligned} \quad (2.35)$$

where the integral is commonly denoted by $\Phi(z)$. Note that $\Phi(z)$ does not involve μ or σ .

Table A3 of Appendix A gives values of $\Phi(z)$ for values of z ranging from 0 to 3.99. Because the normal probability density function is symmetric about the mean, $\Phi(z)$ is evaluated for negative values of z by using the identity $\Phi(-z) = 1 - \Phi(z)$.

Example 2.5 Define a random variable X to be the weight in kilograms of UO_2 powder containers filled at a given facility. Suppose it is known that $X \sim N(22, 0.0016)$, and the cdf $F(22.08) = \Pr(X \leq 22.08)$ is to be evaluated. Applying Equations 2.34 and 2.35 gives $\Pr[Z \leq (22.08 - 22)/0.04] = \Phi(2)$. Entering Table A3 with $z = 2$, we find $\Phi(2) = 0.97725$. This probability is rounded to 0.98 and can be interpreted to mean that 98% of the UO_2 powder containers filled at this facility have weights less than or equal to 22.08 kg.

Suppose now that it is necessary to evaluate the probability that X is between 21.946 and 22.022 kg, that is, $\Pr(21.946 \leq X \leq 22.022)$. Applying Equation 2.34 and Theorem 2.6 gives $\Pr(-1.35 \leq Z \leq 0.55) = \Phi(0.55) - \Phi(-1.35)$. Table A3 gives $\Phi(0.55) = 0.70884$ and $\Phi(-1.35) = 1 - \Phi(1.35) = 1 - 0.91149 = 0.08851$. Thus, $\Phi(0.55) - \Phi(-1.35) = 0.70884 - 0.08851 = 0.62033$ is the desired result. This means that 62% of the UO_2 powder containers filled at the given facility have weights between 21.946 and 22.022 kg.

When computing probabilities from the standard normal distribution, it is helpful to draw sketches that illustrate the necessary steps. Figure 2.4 illustrates the steps in calculating $\Pr(-1.35 \leq Z \leq 0.55)$ in Example 2.5.

Tables of the standard normal distribution are also useful for solving another type of problem. Suppose that in Example 2.5 it is necessary to determine the value of X , say x , such that $\Pr(X \leq x) = 0.90$. From Equation 2.35,

$$\Pr(X \leq x) = \Pr\left(Z \leq \frac{x - 22}{0.04}\right) = \Phi\left(\frac{x - 22}{0.04}\right) = 0.90$$

Entering the body of Table A3, the value of z , for which $\Phi(z) = 0.90$, is between 1.28 and 1.29. For most practical situations, either of these values would suffice, and so $z = 1.28$ is selected. From Equation 2.35,

$$z = \frac{x - 22}{0.04} = 1.28$$

and the solution is

$$x \approx 1.28(0.04) + 22 = 22.0512 \text{ kg}$$

That is, 90% of the UO_2 powder containers filled at the given facility have weights less than or equal to 22.0512 kg. Note that if the value $z = 1.29$ had been chosen, the solution would be $x = 22.0516$ kg.

2.5.2.1 Central Limit Theorem

In numerous material accounting applications, the random variable being considered may be the sum of n independent random variables, some

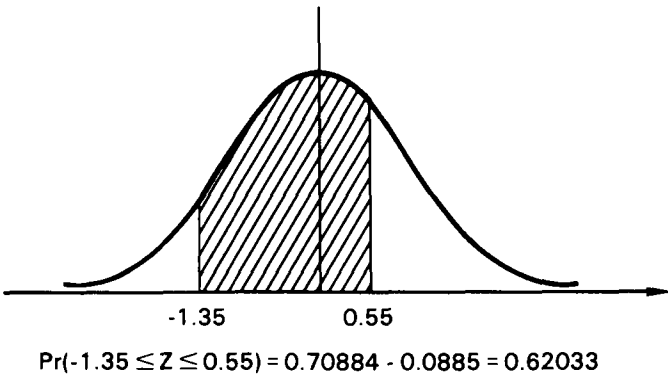
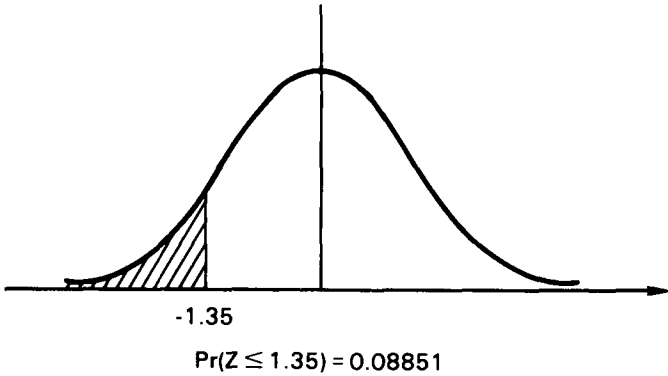
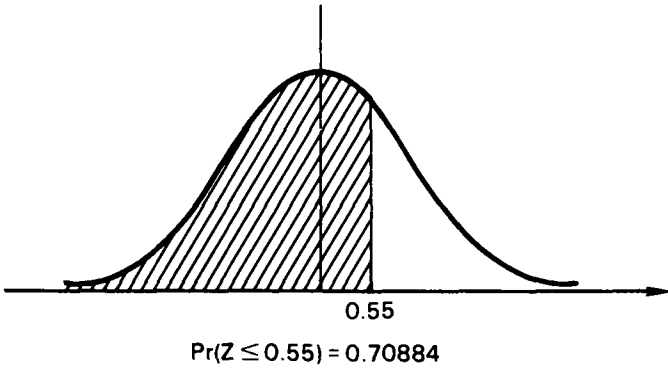


Figure 2.4 Steps in calculating $\Pr(-1.35 \leq Z \leq 0.55)$.

of which may be measurement error, some due to physical considerations, and so forth. A useful result in probability theory states that if a random variable Y is the sum of n independent random variables which satisfy cer-

tain general conditions, then for sufficiently large n , Y is approximately normally distributed. This is known as the **Central Limit Theorem**, which is now stated formally.

Theorem 2.7

If X_1, X_2, \dots, X_n are n independent random variables with $E(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$, and if $Y = X_1 + X_2 + \dots + X_n$, then under some general conditions, the distribution of

$$Z = \frac{Y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad (2.36)$$

approaches a $N(0,1)$ distribution as n approaches ∞ .

The "general conditions" mentioned in the theorem are basically that the terms X_i , taken individually, contribute a negligible amount to the variance of the sum, so that it is unlikely that a single term X_i makes a large contribution to the sum Y . Also, μ_i and σ_i must be finite, but this is not a problem in most applications. The fact that Y is approximately normally distributed when the X_i terms can have essentially any distribution is one of the basic reasons for the usefulness of the normal distribution in statistical inference.

In practical applications of the Central Limit Theorem, the question arises as to how large n must be to obtain valid results. The answer depends on the characteristics of the distribution of the individual X_i terms in Equation 2.36 and the meaning of "valid results." For the special case where the X_i are from the same distribution, as when the X_i are a random sample from a particular distribution, various authors and practitioners advocate some rules of thumb for the size of n when the distribution of X_i falls into one of several categories as follows:

1. **Nearly Normal:** The X_i are from a distribution which is close to a normal distribution in the sense that there is a bell-shaped pdf that is nearly symmetric about the mean. For this case, $n \geq 4$ is adequate (Hines and Montgomery, 1980).

2. **Uniform:** The X_i are from a distribution which has no prominent mode, and the pdf has a rectangular shape which closely imitates the uniform pdf. In this case, $n \geq 12$ has been found to yield satisfactory results (Hines and Montgomery, 1980).

3. **Other Distributions:** When the X_i are from a distribution which does not fall into one of the above two categories, the required size of n depends on the shape of the pdf. For many situations $n = 100$ is not suf-

ficient. This is certainly a nebulous area, and any attempt to give values of n for all specific distributions would be voluminous and probably incomplete. If, however, the X_i are from a distribution with a single mode and a pdf that is not highly skewed, a value of $n \geq 30$ is often sufficient (Hald, 1952).

Also, for the case where the X_i are from a single distribution, Rohatgi (1984) provides a formula for estimating an upper bound on the size of the error that would result from using the normal cdf to approximate the cdf of the sum of the X_i . He also tabulates the error bound estimates as functions of sample size for four specific distributions. Although Rohatgi's formula is valid for any distribution, he points out that the estimate of the error bound can be improved when the cdf of the distribution is completely specified.

It is recommended that if there is any doubt about the assumption that a sum of random variables has a normal distribution, techniques from simulation (Chapter 11) and goodness-of-fit hypothesis testing (Chapter 9) should be implemented to test the validity of the assumption.

2.5.2.2 Normal Approximation to the Binomial

One particularly useful application of the Central Limit Theorem provides a convenient approximation to the binomial distribution. This **normal approximation to the binomial** complements the Poisson approximation discussed in Section 2.4.3. The normal approximation is often appropriate when the Poisson is not and vice versa.

In Section 2.4.2, the binomial random variable X is defined as the number of successes observed in n independent trials of an experiment, where the probability of a success is p at each trial. For the i^{th} single trial, the number of successes, say X_i , is either 1 or 0 with probabilities p and $(1 - p)$, respectively. By substituting $n = 1$ into Equations 2.21 and 2.22, the number of successes X_i at the i^{th} trial has a binomial distribution with mean p and variance $p(1 - p)$. Because the n trials are independent, the random variable X is the sum of n independent random variables X_i , and the Central Limit Theorem is applicable. Using Equation 2.36, Theorem 2.7 states that the random variable Z , given by

$$Z = \frac{X - \sum_{i=1}^n p}{\sqrt{\sum_{i=1}^n p(1-p)}} = \frac{X - np}{\sqrt{np(1-p)}} \quad (2.37)$$

approaches the $N(0,1)$ distribution as n approaches ∞ . Another way of

stating this result is that, for large n , X has an approximate normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1 - p)$.

There are some general rules of thumb concerning how large n must be before the approximation gives reasonably accurate results. In general, the approximation seems to be quite accurate if $np > 5$ when $p \leq 0.5$, or $n(1 - p) > 5$ when $p > 0.5$. This gives $n > 5/p$ when $p \leq 0.5$, and $n > 5/(1 - p)$ when $p > 0.5$. See, for example, Hines and Montgomery (1980).

One major difference between the binomial distribution and the normal distribution is that a binomial random variable is discrete, so that the binomial distribution has probability greater than zero only at $X = 0, 1, \dots, n$, whereas the normal distribution has probability greater than zero in any subinterval on the interval $-\infty \leq X \leq \infty$. This raises a slight problem in applying the normal approximation to the binomial.

Suppose that the binomial probability $\Pr(X = x)$ is to be approximated. The area under the normal pdf is zero at any single value of x . Thus, to accommodate the normal approximation, an interval must be defined which includes x . The simplest way of doing this is to define an interval from $x - 0.5$ to $x + 0.5$ and then use the normal distribution with $\mu = np$ and $\sigma^2 = np(1 - p)$ to compute

$$\begin{aligned} \Pr(X = x) &\cong \Pr(x - 0.5 \leq X \leq x + 0.5) \\ &= \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1 - p)}}\right) - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1 - p)}}\right) \end{aligned} \quad (2.38)$$

Example 2.6 Consider Figure 2.5. The solid vertical lines show the binomial pdf for $n = 10$ and $p = 0.5$. The solid curve is the normal pdf with mean $np = 5$ and variance $np(1 - p) = 2.5$. The rectangles (dotted lines) define a relative frequency histogram where each class has width equal to 1, with midpoint x and class boundaries $x - 0.5$ and $x + 0.5$.

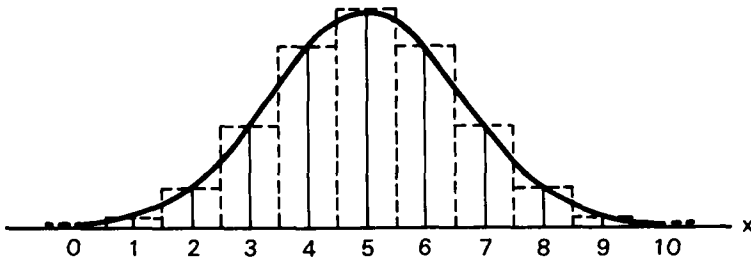


Figure 2.5 Normal approximation to the binomial for $n = 10$ and $p = 0.5$.

The area of each rectangle can be quite accurately approximated by using Equation 2.38. For example,

$$\begin{aligned}\Pr(X = 5) &\cong \Pr(4.5 \leq X \leq 5.5) = \Pr(-0.32 \leq Z \leq 0.32) \\ &= \Phi(0.32) - \Phi(-0.32) = 0.25102\end{aligned}$$

The exact binomial probability, computed from Equation 2.20 is $\Pr(X = 5) = 0.24609$. The difference between the approximation and the exact value is 0.00493, which is negligible for most practical applications.

Cumulative binomial probabilities are formally defined as

$$\Pr(X \leq a) = \sum_{x=0}^a \Pr(X = x)$$

Rather than summing the approximate probabilities to obtain an approximation, use the formula

$$\Pr(X \leq a) \cong \Pr(X \leq a + 0.5) = \Phi\left(\frac{a + 0.5 - np}{\sqrt{np(1-p)}}\right) \quad (2.39)$$

In Example 2.6, where $n = 10$ and $p = 0.5$, suppose that $\Pr(X \leq 7)$ is to be approximated. Using Equation 2.39,

$$\Pr(X \leq 7) \cong \Pr(X \leq 7.5) = \Phi(1.58) = 0.94295$$

The exact probability obtained from Table A1 is 0.9453.

The approximations computed for the previous example are quite close to the exact probabilities. If n were larger than 10 with $p = 0.5$, however, the approximations would be even closer to the exact values.

2.5.3 Lognormal Distribution

The **lognormal distribution** is the distribution of a random variable whose logarithm has a normal distribution. As judged by the number and variety of applications, the lognormal distribution may be as fundamental as the normal distribution in applications of statistics. It applies when random variables are combined by multiplication (rather than addition).

To present the lognormal, consider a random variable X which can take on values in the interval $0 < x < \infty$, where $Y = \log_e X$ is normally distributed with mean μ_Y and variance σ_Y^2 . The quantities μ_Y and σ_Y^2 are the parameters of the lognormal random variable X . The pdf of X is given by

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\{-1/2 [(\log_e x - \mu_Y)/\sigma_Y]^2\}, \quad x > 0 \quad (2.40)$$

A lognormal pdf is illustrated in Figure 2.6. In terms of its parameters, the characteristics of the lognormal distribution can be expressed as follows:

$$\text{mean} = E[X] = \mu_X = \exp(\mu_Y + \sigma_Y^2/2) \quad (2.41)$$

$$\text{variance} = \text{Var}[X] = \sigma_X^2 = [\exp(\sigma_Y^2) - 1] \exp(2\mu_Y + \sigma_Y^2) \quad (2.42)$$

$$\text{mode} = \exp(\mu_Y - \sigma_Y^2) \quad (2.43)$$

$$\text{median} = \exp(\mu_Y) \quad (2.44)$$

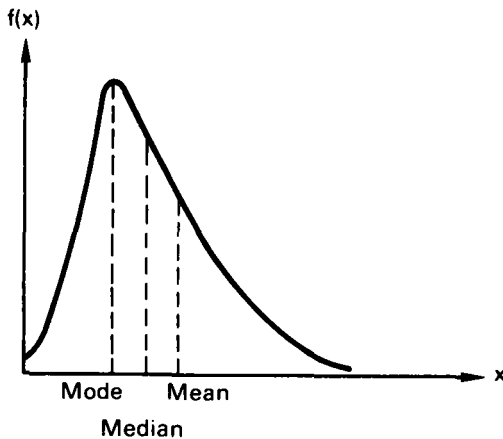


Figure 2.6 The lognormal pdf.

The lognormal distribution has the following useful property: If X_1, X_2, \dots, X_n are n independent random variables from lognormal distributions with parameters μ_{Y_i} and $\sigma_{Y_i}^2$, $i = 1, 2, \dots, n$, and if b, a_1, a_2, \dots, a_n are arbitrary constants, then

$$W = [\exp(b)] \prod_{i=1}^n X_i^{a_i} \quad (2.45)$$

has a lognormal distribution with parameters

$$P_1 = b + \sum_{i=1}^n a_i \mu_{Y_i} \quad (2.46)$$

and

$$P_2 = \sum_{i=1}^n a_i \sigma_{Y_i}^2 \quad (2.47)$$

Thus, the mean and variance of W are given by

$$E[W] = \exp(P_1 + P_2/2)$$

and

$$\text{Var}[W] = [\exp(P_2) - 1] \exp(2P_1 + P_2)$$

A special case of this property is worth noting. If X_1, X_2, \dots, X_n are independent random variables having a lognormal distribution with parameters μ_Y and σ_Y^2 , then the geometric mean, given in Equation 1.5 by

$$G = \left(\prod_{i=1}^n X_i \right)^{1/n} \quad (2.48)$$

has a lognormal distribution with parameters μ_Y and σ_Y^2/n . Thus, the geometric mean G has mean and variance given by

$$E[G] = \exp(\mu_Y + \sigma_Y^2/2n) \quad (2.49)$$

and

$$\text{Var}[G] = [\exp(\sigma_Y^2/n) - 1] \exp(2\mu_Y + \sigma_Y^2/n) \quad (2.50)$$

The popularity of the lognormal is due in part to the convenience of working in terms of $Y = \log_e X$, where methods based on the normal distribution are easily applied.

Example 2.7 Suppose that the random variable $Y = \log_e X$ has an $N(3,2)$ distribution. Then X has a lognormal distribution with

$$E[X] = \exp(3 + 2/2) = 54.60$$

$$\text{Var}[X] = [\exp(2) - 1] \exp(6 + 2) = 19,045.51$$

$$\text{mode} = \exp(3 - 2) = 2.72$$

$$\text{median} = \exp(3) = 20.09$$

The probability that $X \leq 100$, for example, can be expressed in terms of $Y = \log_e X$ and determined from Table A3 as follows:

$$\begin{aligned} \Pr(X \leq 100) &= \Pr(Y \leq \log_e 100) \\ &= \Pr\left[Z \leq \frac{4.61 - 3}{\sqrt{2}}\right] \\ &= \Phi(1.14) = 0.87285 \end{aligned}$$

2.5.4 Other Continuous Distributions

There are a large number of both discrete and continuous distributions in statistical theory that are useful in many specific areas of application. Those presented in the previous sections of this chapter are among the more commonly used and applicable distributions in nuclear material accounting. Also, some commonly used sampling distributions are presented in Chapter 3 and in subsequent chapters as needed. In this section, a brief introduction is given to four continuous distributions which have the potential of being used for some nuclear material accounting or related applications. These are the **exponential**, **gamma**, **Weibull**, and **beta** distributions. For a detailed presentation of these and other distributions, the reader is referred to Johnson and Kotz (1969; 1970) or Hahn and Shapiro (1967).

The **exponential distribution** has pdf, parameter, mean, and variance as given in Table 2.5. The exponential is used to model the time (or space, distance, etc.) between occurrences of events and is related to the Poisson distribution as follows: If the time t between occurrences of an event is a random variable that has an exponential distribution with parameter λ , then the number of events occurring in an interval of length T has a Poisson distribution, as given by Equation 2.25, with parameter $\mu = \lambda T$.

The **gamma distribution** has pdf, parameters, mean, and variance as given in Table 2.5. The parameter r is called the **shape parameter**, and λ

TABLE 2.5
Summary of Four Continuous Distributions

Distribution	Parameters	pdf	Mean	Variance
Exponential	$\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$	$1/\lambda$	$1/\lambda^2$
Gamma	$r > 0$ $\lambda > 0$	$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$	r/λ	r/λ^2
Weibull	$-\infty < \gamma < \infty$ $\delta > 0$ $\beta > 0$	$f(x) = \begin{cases} \frac{\beta}{\delta} \left[\frac{x - \gamma}{\delta} \right]^{\beta-1} \exp \left[- \left[\frac{x - \gamma}{\delta} \right]^\beta \right], & x \geq \gamma \\ 0, & \text{otherwise} \end{cases}$	$\gamma + \delta \Gamma \left[\frac{1}{\beta} + 1 \right]$	$\delta^2 \left\{ \Gamma \left[\frac{2}{\beta} + 1 \right] - \left[\Gamma \left[\frac{1}{\beta} + 1 \right] \right]^2 \right\}$
Beta	$\alpha > 0$ $\beta > 0$	$f(x) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

is called the **scale parameter**. There is a close relationship between the gamma and the exponential distributions, as follows: If a random variable X is the sum of r independent, exponentially distributed random variables, each having parameter λ , then X has a gamma distribution with parameters r and λ . Note that when $r = 1$, the gamma and exponential pdf's are identical.

A special form of the gamma distribution, known as the **chi-square distribution**, arises when a random variable X is the sum of the squared values of n independent random variables from the $N(0,1)$ distribution. In this case the parameters are $\lambda = 1/2$ and $r = n/2$, and X is said to have a chi-square distribution with n **degrees of freedom**. The chi-square distribution and some applications are discussed further in subsequent chapters.

The **Weibull distribution**, as specified in Table 2.5, has been widely applied in modeling many types of random phenomena. The popularity of the Weibull distribution stems from the fact that it provides excellent approximations to many different distributions. Also, the Weibull pdf is easily integrated yielding a simple form for the cdf. Thus, no tables are necessary for its application. Its parameters are classified as the **location parameter** γ , the **scale parameter** δ , and the **shape parameter** β . If values of these parameters are appropriately chosen, the Weibull pdf will closely approximate many observational phenomena. Note that when $\gamma = 0$ and $\beta = 1$, the Weibull pdf is identical to an exponential pdf with parameter $\lambda = 1/\delta$.

The **beta distribution**, as specified in Table 2.5, has a wide variety of applications. It is used in conjunction with order statistics for constructing confidence intervals for distribution quantiles and tolerance limits for distribution functions. A detailed discussion of this application with examples is given by Hogg and Craig (1978). The beta pdf also occurs frequently as both a prior and a posterior pdf in deriving so-called Bayes estimators, especially for percentages and binomial probabilities. An excellent account of this application is given by Larson (1969). Note that when $\alpha = \beta = 1$ the beta reduces to a uniform distribution on the unit interval.

2.6 JOINTLY DISTRIBUTED RANDOM VARIABLES

The presentation thus far has been concerned with distributions which involve only one random variable. Such distributions are referred to as **univariate** distributions. Situations often arise in practice, however, where more than one random variable is measured on each sampled unit. In such cases, the distribution of the random variables is called a **multivariate** distribution.

This section presents concepts, definitions, and formulas that are used when working with more than one random variable at a time. To illustrate the concepts, the special case of two random variables (the **bivariate case**) is considered, first with both variables discrete and then with both variables continuous. A generalization to more than two random variables is then given, and two commonly used distributions, the multinomial and the bivariate normal, are introduced.

2.6.1 Discrete Bivariate Distributions

If X and Y are two discrete random variables, the **joint probability density function (pdf)** is defined as

$$f(x,y) = \Pr(X = x, Y = y) \quad (2.51)$$

where $f(x,y) \geq 0$ and $\sum_x \sum_y f(x,y) = 1$.

Suppose that X can assume any one of m values x_1, x_2, \dots, x_m , and Y can assume any one of n values y_1, y_2, \dots, y_n . Then the probability that $X = x_i$ and $Y = y_k$ is given by

$$\Pr(X = x_i, Y = y_k) = f(x_i, y_k) \quad (2.52)$$

This joint pdf of X and Y can be represented by a joint probability table, such as Table 2.6. The probability that $X = x_i$ is obtained by adding all entries in the row corresponding to x_i and is given by

$$f_1(x_i) = \Pr(X = x_i) = \sum_{k=1}^n f(x_i, y_k) \quad (2.53)$$

TABLE 2.6
Joint and Marginal pdf's of X and Y

$X \backslash Y$	y_1	y_2	\dots	y_n	Totals
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	\dots	$f(x_1, y_n)$	$f_1(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	\dots	$f(x_2, y_n)$	$f_1(x_2)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	\dots	$f(x_m, y_n)$	$f_1(x_m)$
Totals	$f_2(y_1)$	$f_2(y_2)$	\dots	$f_2(y_n)$	1

For $i = 1, 2, \dots, m$, the values of $f_1(x_i)$ are given by the row totals in the right-hand margin (last column) of Table 2.6. Similarly, the probability that $Y = y_k$ is obtained by summing all entries in the column corresponding to y_k and is given by

$$f_2(y_k) = P(Y = y_k) = \sum_{i=1}^m f(x_i, y_k) \quad (2.54)$$

For $k = 1, 2, \dots, n$, the values of $f_2(y_k)$ are given by the column totals in the lower margin (bottom row) of Table 2.6.

Because $f_1(x_i)$ and $f_2(y_k)$ are obtained from the margins of the table, they are commonly called the **marginal pdf's** of X and Y , respectively, and are generally denoted by $f_1(x)$ and $f_2(y)$. It should be noted that marginal pdf's have the same properties as univariate pdf's in the sense that $f_1(x) \geq 0$, $f_2(y) \geq 0$, and

$$\sum_x f_1(x) = \sum_y f_2(y) = \sum_x \sum_y f(x, y) = 1$$

The **joint cumulative distribution function (cdf)** of X and Y is given by

$$F(x, y) = \Pr(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v) \quad (2.55)$$

To compute $F(x, y)$ from Table 2.6, sum over all values of $f(x_i, y_k)$ for which $x_i \leq x$ and $y_k \leq y$.

Suppose now that only Y is measured, but it is of interest to determine the probability that $X = x_i$ when it is known that $Y = y_k$. Such probabilities are given by the **conditional pdf** of X , given $Y = y$, denoted by $f(x|y)$, which is determined from Definition 2.11 to be

$$f(x|y) = \frac{f(x, y)}{f_2(y)} \quad (2.56)$$

For $x = x_i$ and $y = y_k$, the values of $f(x_i|y_k)$ are computed from Table 2.6 by dividing the elements of the y_k column by the marginal probability, $f_2(y_k)$.

Similarly the conditional pdf of Y , given $X = x$, is given by

$$f(y|x) = \frac{f(x, y)}{f_1(x)} \quad (2.57)$$

For $x = x_i$ and $y = y_k$, the values of $f(y_k|x_i)$ are computed from Table 2.6 by dividing the elements of the x_i row by the marginal probability $f_1(x_i)$.

Conditional pdf's have the same properties as univariate pdf's in the sense that $f(x|y) \geq 0$ and

$$\sum_x f(x|y) = \frac{\sum_x f(x,y)}{f_2(y)} = \frac{f_2(y)}{f_2(y)} = 1$$

If the conditional pdf $f(x|y)$ does not involve y , then X is independent of Y in the probability sense (Definition 2.12). When this is the case, it can be shown that

$$f(x,y) = f_1(x)f_2(y) \quad (2.58)$$

Thus, when X and Y are independent,

$$f(x|y) = \frac{f_1(x)f_2(y)}{f_2(y)} = f_1(x), \text{ for all } y \quad (2.59)$$

and

$$f(y|x) = \frac{f_1(x)f_2(y)}{f_1(x)} = f_2(y), \text{ for all } x \quad (2.60)$$

Expected values are obtained by extending the concepts presented in the previous sections. The mean of X is given by

$$\mu_X = E[X] = \sum_x xf_1(x) = \sum_x \sum_y xf(x,y) \quad (2.61)$$

Similarly, the mean of Y is given by

$$\mu_Y = E[Y] = \sum_y yf_2(y) = \sum_x \sum_y yf(x,y) \quad (2.62)$$

The variance of X is given by

$$\sigma_X^2 = \text{Var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2 = \sum_x x^2 f_1(x) - \mu_X^2 \quad (2.63)$$

Similarly, the variance of Y is given by

$$\sigma_Y^2 = \text{Var}[Y] = E[(Y - \mu_Y)^2] = E[Y^2] - \mu_Y^2 = \sum_y y^2 f_2(y) - \mu_Y^2 \quad (2.64)$$

An expected value that is of great interest in most bivariate applications is the **covariance** of X and Y , which is given by

$$\begin{aligned}\sigma_{XY} &= \text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X\mu_Y \\ &= \sum_x \sum_y xyf(x, y) - \mu_X\mu_Y\end{aligned}\quad (2.65)$$

A measure of association between X and Y , known as the **correlation coefficient**, is defined as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\quad (2.66)$$

Note that when X and Y are independent, $E[XY] = \mu_X\mu_Y$, and thus $\sigma_{XY} = 0$ and $\rho_{XY} = 0$. The converse is not always true, however. That is, in general, $\rho_{XY} = 0$ does not imply independence.

The conditional mean and variance of X , given $Y = y$, are defined as

$$\mu_{X|y} = E[X|y] = \sum_x xf(x|y)\quad (2.67)$$

and

$$\sigma_{X|y}^2 = \text{Var}[X|y] = \sum_x x^2f(x|y) - \mu_{X|y}^2\quad (2.68)$$

Similarly, the conditional mean and variance of Y , given $X = x$, are defined as

$$\mu_{Y|x} = E[Y|x] = \sum_y yf(y|x)\quad (2.69)$$

and

$$\sigma_{Y|x}^2 = \text{Var}[Y|x] = \sum_y y^2f(y|x) - \mu_{Y|x}^2\quad (2.70)$$

2.6.2 Continuous Bivariate Distributions

For the case where two random variables X and Y are both continuous, practically all of the definitions can be obtained by analogy to the discrete case if the summations are replaced by integrals.

If X and Y are continuous random variables, any function $f(x, y)$ may be regarded as a **joint pdf** of X and Y if

$$f(x, y) \geq 0$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$$

The **marginal pdf's** of X and Y are given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y) dy \quad (2.71)$$

and

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y) dx \quad (2.72)$$

It should be noted that the marginal pdf's have the same properties as univariate pdf's in the sense that $f_1(x) \geq 0$, $f_2(y) \geq 0$, and

$$\int_{-\infty}^{\infty} f_1(x) dx = \int_{-\infty}^{\infty} f_2(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$$

The **joint cumulative distribution function** of X and Y is given by

$$F(x,y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(u,v) du dv \quad (2.73)$$

and by analogy with Equation 2.4, the joint pdf of X and Y is defined as

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \quad (2.74)$$

The **conditional pdf** of X, given $Y = y$, is defined as

$$f(x|y) = \frac{f(x,y)}{f_2(y)} \quad (2.75)$$

and the conditional pdf of Y, given $X = x$, is defined as

$$f(y|x) = \frac{f(x,y)}{f_1(x)} \quad (2.76)$$

Conditional pdf's have the same properties as univariate pdf's in the sense that $f(x|y) \geq 0$, $f(y|x) \geq 0$,

$$\int_{-\infty}^{\infty} f(x|y)dx = 1$$

and

$$\int_{-\infty}^{\infty} f(y|x)dy = 1$$

If the conditional pdf $f(x|y)$ does not involve y , then X and Y are said to be **independent** in the probability sense (Definition 2.12). When this is the case, it can be shown that

$$f(x,y) = f_1(x)f_2(y) \quad (2.77)$$

Thus, when X and Y are independent,

$$f(x|y) = \frac{f_1(x)f_2(y)}{f_2(y)} = f_1(x), \text{ for all } y \quad (2.78)$$

and

$$f(y|x) = \frac{f_1(x)f_2(y)}{f_1(x)} = f_2(y), \text{ for all } x \quad (2.79)$$

Expected values are obtained by analogy with Equations 2.60 through 2.70 if the summations are replaced by integrals. These are listed here.

$$\mu_X = E[X] = \int_{-\infty}^{\infty} xf_1(x)dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)dxdy \quad (2.80)$$

$$\mu_Y = E[Y] = \int_{-\infty}^{\infty} yf_2(y)dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dxdy \quad (2.81)$$

$$\sigma_X^2 = \text{Var}[X] = \int_{-\infty}^{\infty} x^2f_1(x)dx - \mu_X^2 \quad (2.82)$$

$$\sigma_Y^2 = \text{Var}[Y] = \int_{-\infty}^{\infty} y^2f_2(y)dy - \mu_Y^2 \quad (2.83)$$

$$\sigma_{XY} = E[XY] - \mu_X\mu_Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy - \mu_X\mu_Y \quad (2.84)$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2.85)$$

Note that when X and Y are independent, $E[XY] = \mu_X\mu_Y$, and thus $\sigma_{XY} = 0$ and $\rho_{XY} = 0$. As previously noted for the discrete case, $\rho_{XY} = 0$ does not, in general, imply independence. For normal distributions, however, $\rho_{XY} = 0$ does imply independence.

The conditional means and variances are given by

$$\mu_{X|Y} = \int_{-\infty}^{\infty} xf(x|y)dx \quad (2.86)$$

$$\mu_{Y|X} = \int_{-\infty}^{\infty} yf(y|x)dy \quad (2.87)$$

$$\sigma_{X|Y}^2 = \int_{-\infty}^{\infty} x^2f(x|y)dx - \mu_{X|Y}^2 \quad (2.88)$$

and

$$\sigma_{Y|X}^2 = \int_{-\infty}^{\infty} y^2f(y|x)dy - \mu_{Y|X}^2 \quad (2.89)$$

2.6.3 More Than Two Variables

The concepts presented in Sections 2.6.1 and 2.6.2 are easily extended to the general case of k jointly distributed random variables, say X_1, X_2, \dots, X_k . In general, any function f may be regarded as a joint pdf of k random variables if

$$f(x_1, x_2, \dots, x_k) \geq 0$$

and

$$\sum_{x_1} \sum_{x_2} \dots \sum_{x_k} f(x_1, x_2, \dots, x_k) = 1$$

when all X_i are discrete, or

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k = 1$$

when all X_i are continuous.

For k jointly distributed random variables, the joint cdf is given by

$$F(x_1, x_2, \dots, x_k) = \sum_{u_1 \leq x_1} \sum_{u_2 \leq x_2} \dots \sum_{u_k \leq x_k} f(u_1, u_2, \dots, u_k) \quad (2.90)$$

for the discrete case, and

$$F(x_1, x_2, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(u_1, u_2, \dots, u_k) du_1 du_2 \dots du_k \quad (2.91)$$

for the continuous case. Thus, for the continuous case, the joint pdf can be expressed as

$$f(x_1, x_2, \dots, x_k) = \frac{\partial^k F(x_1, x_2, \dots, x_k)}{\partial x_1 \partial x_2 \dots \partial x_k} \quad (2.92)$$

The **joint marginal pdf** of any subset of the k random variables is obtained by summing (or integrating) the joint pdf $f(x_1, x_2, \dots, x_k)$ over the entire range of all other random variables which are not in the subset. For example, the joint marginal pdf of X_1, X_2 , and X_3 is given by

$$f_{123}(x_1, x_2, x_3) = \sum_{x_4} \sum_{x_5} \dots \sum_{x_k} f(x_1, x_2, \dots, x_k) \quad (2.93)$$

for the discrete case, and

$$f_{123}(x_1, x_2, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k) dx_4 dx_5 \dots dx_k \quad (2.94)$$

for the continuous case.

The **conditional joint pdf** of a subset of the random variables, given the values of the other random variables not in the subset, is found by taking the ratio of the joint pdf of all k variables to the joint marginal pdf of the variables not in the subset. For example, the conditional joint pdf of X_4, X_5, \dots, X_k , given that $X_1 = x_1, X_2 = x_2$, and $X_3 = x_3$, is defined as

$$f(x_4, x_5, \dots, x_k | x_1, x_2, x_3) = \frac{f(x_1, x_2, \dots, x_k)}{f_{123}(x_1, x_2, x_3)} \quad (2.95)$$

Similarly, the conditional joint pdf can be determined for a subset of the random variables, given the values of another subset of the random variables, where the two subsets do not exhaust all k random variables. This is accomplished by applying the above rule to the joint marginal pdfs. For example, the conditional pdf of X_1 , given $X_2 = x_2$ and $X_3 = x_3$, has the form

$$f(x_1 | x_2, x_3) = \frac{f_{123}(x_1, x_2, x_3)}{f_{23}(x_2, x_3)} \quad (2.96)$$

The expected values for the general k variable case are summarized using vector and matrix notation. The k element **mean vector** is denoted by

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} \quad (2.97)$$

where the i^{th} element μ_i is defined as

$$\mu_i = E[X_i] = \sum_{x_i} x_i f_i(x_i) \quad (2.98)$$

for the discrete case, and

$$\mu_i = E[X_i] = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i \quad (2.99)$$

for the continuous case. Note that in Equations 2.98 and 2.99, $f_i(x_i)$ is the marginal pdf of x_i obtained by summing (or integrating) $f(x_1, x_2, \dots, x_k)$ over all other variables.

The k -by- k symmetric **variance-covariance matrix** is denoted by

$$\underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2k} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdot & \vdots \\ \vdots & \vdots & \vdots & \cdot & \vdots \\ \vdots & \vdots & \vdots & \cdot & \sigma_{k-1,k} \\ \sigma_{k1} & \sigma_{k2} \dots & \sigma_{k,k-1} & \sigma_k^2 \end{bmatrix} \quad (2.100)$$

where the diagonal elements σ_i^2 are the variances of the X_i , given by

$$\sigma_i^2 = \text{Var}[X_i] = E[X_i^2] - \mu_i^2 = \sum_{x_i} x_i^2 f_i(x_i) - \mu_i^2 \quad (2.101)$$

for the discrete case, and

$$\sigma_i^2 = \text{Var}[X_i] = E[X_i^2] - \mu_i^2 = \int_{-\infty}^{\infty} x_i^2 f(x_i) dx_i - \mu_i^2 \quad (2.102)$$

for the continuous case. The i th off-diagonal element σ_{ij} is the covariance of X_i and X_j , which by analogy with Equations 2.65 and 2.84 is given by

$$\begin{aligned}\sigma_{ij} &= \text{Cov}[X_i, X_j] = E[X_i X_j] - \mu_i \mu_j \\ &= \sum_{x_i} \sum_{x_j} x_i x_j f_{ij}(x_i, x_j) - \mu_i \mu_j\end{aligned}\quad (2.103)$$

for the discrete case, and

$$\begin{aligned}\sigma_{ij} &= \text{Cov}[X_i, X_j] = E[X_i X_j] - \mu_i \mu_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j f_{ij}(x_i, x_j) dx_i dx_j - \mu_i \mu_j\end{aligned}\quad (2.104)$$

for the continuous case.

The correlation ρ_{ij} between X_i and X_j is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}\quad (2.105)$$

If X_i and X_j are independent, then $\sigma_{ij} = 0$, and thus $\rho_{ij} = 0$. In general, however, $\rho_{ij} = 0$ does not imply independence.

A multivariate joint pdf that deserves special mention arises when X_1, X_2, \dots, X_n is a random sample of size n from the pdf $f(x)$. In this case, the X_i are assumed to be independent, so that by analogy with Equations 2.58 and 2.77, the joint pdf is given by

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n) = \prod_{i=1}^n f(x_i)\quad (2.106)$$

This joint pdf is called the **likelihood function** and is discussed further in Chapter 3.

2.6.4 Two Special Multivariate Distributions

In this section, two multivariate distributions are presented that have applications in nuclear material accounting. These are the **multinomial distribution** and the **bivariate normal distribution**.

2.6.4.1 Multinomial Distribution

The **multinomial distribution** is the multivariate extension of the binomial distribution. It is a discrete distribution that arises in attribute sampling applications when the attribute of interest can be classified into $k \geq 2$ mutually exclusive classes. That is, in a sequence of independent

trials, there are $k \geq 2$ possible outcomes at each trial. Let the probabilities of the k outcomes be denoted by p_1, p_2, \dots, p_k , where

$$\sum_{i=1}^k p_i = 1 \quad (2.107)$$

The values of the p_i are assumed constant for all trials.

Consider n independent trials of such an experiment. Let X_i be the number of times the outcome corresponding to p_i occurs. The joint pdf of X_1, X_2, \dots, X_k is given by

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} \quad (2.108)$$

where

$$\sum_{i=1}^k x_i = n \quad (2.109)$$

It should be noted that X_1, X_2, \dots, X_k are not independent random variables because knowing the value of n and the values of any $k - 1$ of the X_i completely determines the value of the remaining X_i . For example, if it is known that $X_1 = x_1, X_2 = x_2, \dots, X_{k-1} = x_{k-1}$, then the value of X_k is determined from Equation 2.109 and is given by

$$X_k = x_k = n - \sum_{i=1}^{k-1} x_i$$

Similarly, knowing the values of any $k - 1$ of the p_i , the value of the remaining p_i can be determined from Equation 2.107. Using the same example as above, if the values of p_1, p_2, \dots, p_{k-1} are known, then the value of p_k is given by

$$p_k = 1 - \sum_{i=1}^{k-1} p_i$$

For this reason, a multinomial distribution with k possible outcomes is often said to be a $k - 1$ variate distribution with parameters $n, p_1, p_2, \dots, p_{k-1}$. The quantity p_k plays the same role as $(1 - p)$ in the bino-

mial distribution. For example, the binomial has $k = 2$ possible outcomes and is classified as a univariate distribution with parameters n and p , where the number of successes is the random variable.

The mean and variance of each X_i are given by

$$\mu_i = E[X_i] = np_i \quad (2.110)$$

and

$$\sigma_i^2 = \text{Var}[X_i] = np_i(1 - p_i) \quad (2.111)$$

The covariance σ_{ij} between any two variables X_i and X_j , for $i \neq j$, is given by

$$\sigma_{ij} = \text{Cov}[X_i, X_j] = -np_i p_j, \quad i \neq j \quad (2.112)$$

Example 2.8 Consider an automated process where UO_2 powder containers are filled to a prespecified weight, sealed, and labeled. The plant manager routinely audits the container weights, which are rounded to the nearest 10 g, to monitor their compliance with the prespecified filled weight. Historical audit data reveal that when the process is “in control,” 95% of the container weights will agree exactly with the prespecified filled weight, 4% will agree within 10 g, and 1% will have discrepancies of more than 10 g. It is desirable that these percentages remain constant over time.

An audit proceeds by randomly selecting 50 filled containers from the process, weighing them, and recording their weights to the nearest 10 g. Let

X_1 = number of containers whose weights agree exactly with the prespecified filled weight

X_2 = number of containers with weight discrepancies within 10 g

X_3 = number of containers with weight discrepancies exceeding 10 g

The process is said to be “in control” if, in the sample of 50 containers, values of $X_2 \leq 3$ and $X_3 \leq 1$ are observed. Suppose that it is of interest to compute the probability $\text{Pr}(X_2 \leq 3, X_3 \leq 1)$ when the percentages are as stated above.

The appropriate pdf for computing probabilities is the multinomial, given by

$$f(x_1, x_2, x_3) = \frac{50!}{x_1! x_2! x_3!} (0.95)^{x_1} (0.04)^{x_2} (0.01)^{x_3}$$

The desired probability is computed as

$$\begin{aligned} \Pr(X_2 \leq 3, X_3 \leq 1) &= f(50,0,0) + f(49,0,1) + f(49,1,0) + f(48,1,1) \\ &\quad + f(48,2,0) + f(47,2,1) + f(47,3,0) + f(46,3,1) \\ &= 0.7828 \end{aligned}$$

where, for example,

$$f(48,2,0) = \frac{50!}{48!2!0!} (0.95)^{48} (0.04)^2 (0.01)^0 = 0.16710$$

Thus, when the percentages are as specified above, there is a 78% chance of concluding that the process is in control and a 22% chance of concluding that the process is out of control. It would seem desirable to develop an inspection plan which would yield a larger probability of concluding that the process is in control when the percentages are at the desired levels. This could be accomplished by increasing the sample size n .

For this example, the mean vector is computed from Equation 2.110 and is given by

$$\underline{\mu} = \begin{bmatrix} 47.5 \\ 2.0 \\ 0.5 \end{bmatrix}$$

These are the average values we would expect to observe if the sampling process were repeated a large number of times when the percentages are at the desired levels.

The variance-covariance matrix for this example is computed from Equations 2.111 and 2.112 and is given by

$$\underline{\Sigma} = \begin{bmatrix} 2.375 & -1.900 & -0.475 \\ -1.900 & 1.920 & -0.020 \\ -0.475 & -0.020 & 0.495 \end{bmatrix}$$

2.6.4.2 Bivariate Normal Distribution

One of the most important and widely used multivariate distributions in statistics is the **multivariate normal distribution**. This distribution and its

various applications have been the topic of many excellent texts on multivariate analysis, such as Anderson (1984) and Kshirsagar (1972). To discuss more than two variables, however, is quite tedious without the use of matrices and vectors. The two variable case, the **bivariate normal distribution**, is presented in this section.

Two jointly distributed random variables X_1 and X_2 have a **bivariate normal distribution** if their joint pdf is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}, \quad (2.113)$$

$$-\infty < x_1 < \infty \text{ and } -\infty < x_2 < \infty$$

The parameters μ_1 , μ_2 , σ_1 , σ_2 , and ρ have straightforward interpretations as the means, standard deviations, and correlation coefficient of X_1 and X_2 . That is,

$$\begin{aligned} E[X_1] &= \mu_1 \\ E[X_2] &= \mu_2 \\ \text{Var}[X_1] &= \sigma_1^2 \\ \text{Var}[X_2] &= \sigma_2^2 \\ \rho &= \rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2} \end{aligned}$$

The joint probability that $a < X_1 < b$ and $c < X_2 < d$ is computed from the formula

$$\Pr(a < X_1 < b, c < X_2 < d) = \int_c^d \int_a^b f(x_1, x_2) dx_1 dx_2 \quad (2.114)$$

and is illustrated in Figure 2.7.

The marginal pdf's of X_1 and X_2 are given by

$$f_1(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp \left\{ -\frac{(x_1 - \mu_1)^2}{2\sigma_1^2} \right\} \quad (2.115)$$

and

$$f_2(x_2) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left\{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}\right\} \quad (2.116)$$

On comparison with Equation 2.32, note that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. That is, the marginal pdf's of X_1 and X_2 are univariate normal pdf's.

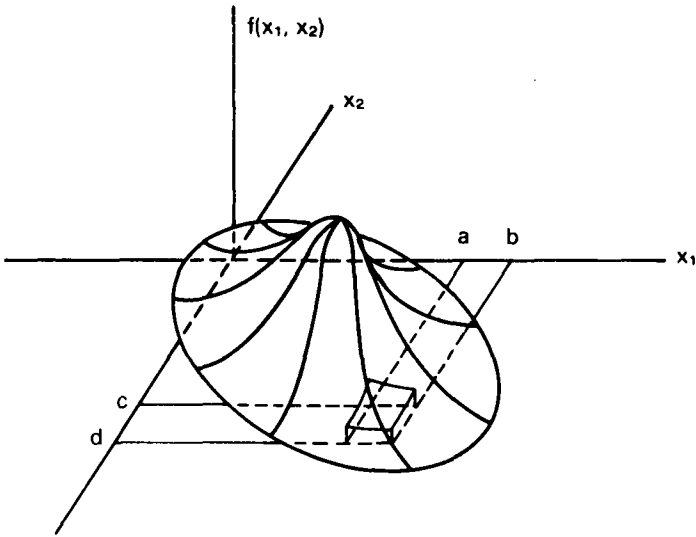


Figure 2.7 The bivariate normal pdf.

An important application of the bivariate normal distribution in nuclear material accounting involves finding the conditional distribution of one variable given the value of the other. If X_1 and X_2 have a joint pdf $f(x_1, x_2)$, given by Equation 2.113, then in accordance with Equation 2.75 the conditional pdf of X_1 , given that $X_2 = x_2$, is given by

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1 - \mu_1 - \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2)\right]^2\right\} \quad (2.117)$$

where $f_2(x_2)$ is given by Equation 2.116. On comparison with Equation 2.32, $f(x_1|x_2)$ is a univariate normal pdf with mean and variance given by

$$E[X_1|x_2] = \mu_1 + \rho(\sigma_1/\sigma_2)(x_2 - \mu_2) \quad (2.118)$$

and

$$\text{Var}[X_1|x_2] = \sigma_1^2(1 - \rho^2) \quad (2.119)$$

Note that $\text{Var}[X_1|x_2]$ is constant for all values of X_2 .

The conditional pdf of X_2 , given that $X_1 = x_1$, and its mean and variance are determined by exchanging the subscripts 1 and 2 in Equations 2.117, 2.118, and 2.119.

It should be noted with the bivariate normal distribution that if $\rho = 0$, $f(x_1|x_2)$ reduces to $f_1(x_1)$. That is, Equation 2.117 reduces to Equation 2.115, from which it follows that $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. Thus, when $\rho = 0$, X_1 and X_2 are independent.

The expression for $E[X_1|x_2]$ is commonly called the **regression** of X_1 on X_2 and is usually written as

$$E[X_1|x_2] = \beta_0 + \beta_1 x_2 \quad (2.120)$$

where

$$\beta_0 = \mu_1 - \beta_1 \mu_2$$

and

$$\beta_1 = \rho \sigma_1 / \sigma_2$$

In many applications it is desirable to control one of the variables, e.g., X_2 , and only allow it to take on certain values. These values are called **fixed levels** of X_2 . At each fixed level of X_2 , a random sample of values of X_1 is taken. The analysis of data from this type of controlled experiment requires special assumptions. Regression analysis is the topic of Chapter 6, and the generalizations and underlying assumptions for various applications are presented there.

Example 2.9 Use of a nondestructive assay (NDA) instrument is being considered as an alternative to a more expensive but very reliable procedure for measuring the plutonium content of specific items. Over an extended period of time, both techniques were used simultaneously to

measure a large number of items. Thus, for each item, values were recorded for the two random variables:

X_1 = grams of plutonium as determined by the more expensive procedure

X_2 = net count rate (counts per minute) from the NDA instrument, corrected for background and dead time

It was concluded that X_1 and X_2 have a bivariate normal pdf with parameters:

$$\begin{aligned}\mu_1 &= 290 \text{ g} \\ \mu_2 &= 390,217 \text{ cpm} \\ \sigma_1 &= 34 \text{ g} \\ \sigma_2 &= 44,307 \text{ cpm} \\ \rho &= 0.99\end{aligned}$$

Note that σ_1 primarily reflects the variability in plutonium content among the measured items and is not an indicator of the precision of the more expensive procedure. Because the correlation coefficient is 0.99, the NDA instrument could replace the more expensive procedure. It is of interest to determine the conditional pdf of X_1 when the value of X_2 is known. If the above parameters are substituted into Equations 2.117, 2.118, and 2.105, the conditional pdf $f(x_1|x_2)$ is a univariate normal pdf with mean and variance given by

$$\begin{aligned}E[X_1|x_2] &= 290 + 0.99 \left[\frac{34}{44,307} \right] (x_2 - 390,217) \\ &= -6.4476 + x_2(0.759699)10^{-3} \text{ g}\end{aligned}$$

and

$$\text{Var}[X_1|x_2] = (34)^2(1 - 0.99^2) = 23.0044 \text{ g}^2$$

Suppose that an unknown item is counted using the NDA instrument, and its net corrected count rate is $x_2 = 350,000$ cpm. Then the amount of plutonium contained in this item is from a normal distribution with mean and variance given by

$$E[X_1|x_2 = 350,000] = 259.45 \text{ g}$$

and

$$\text{Var}[X_1|x_2 = 350,000] = 23.0044 \text{ g}^2$$

It is of interest to find an upper limit (U) and a lower limit (L) such that

$$\Pr(X_1 > U | x_2 = 350,000) = 0.05$$

and

$$\Pr(X_1 < L | x_2 = 350,000) = 0.05$$

Using the procedures discussed in Section 2.5.2,

$$\Pr[X_1 > (U - 259.45)/\sqrt{23.0044}] = 1 - \Phi\left(\frac{U - 259.45}{4.80}\right) = 0.05$$

or

$$\Phi\left(\frac{U - 259.45}{4.80}\right) = 0.95$$

Entering Table A3, the value of z for which $\Phi(z) = 0.95$ is $z = 1.65$. Thus,

$$\frac{U - 259.45}{4.80} = 1.65$$

or

$$U = 267.36 \text{ g}$$

Similarly, the value of z for which $\Phi(z) = 0.05$ is $z = -1.65$, so that

$$\frac{L - 259.45}{4.80} = -1.65$$

or

$$L = 251.54 \text{ g}$$

Thus, the following probability statement can be made about the amount of plutonium in the item:

$$\Pr(251.54 \text{ g} < X_1 < 267.36 \text{ g}) = \Pr(-1.65 < Z < 1.65) = 0.90$$

That is, the probability is 0.90 that the amount of plutonium in the item is between 251.54 g and 267.36 g.

This example is a problem in **calibration**, a topic which is more thoroughly developed in Chapter 15.

2.7 FUNCTIONS OF RANDOM VARIABLES

In nuclear material control and accounting applications, the random variable of interest is often a mathematical function of other random variables, and it is desirable to characterize the pdf of this composite variable. Depending upon the form of the mathematical function and whether or not the pdf's of the individual random variables in the function are known, it may be difficult or impossible to derive its pdf. In most cases, however, either exact or approximate formulas for the mean and variance of the function can be derived. In most applications, this provides an adequate characterization of the pdf for estimation and inferential purposes.

Section 2.7.1 presents the exact formulas for the mean and variance of a linear combination of random variables. Section 2.7.2 presents a method for deriving approximate formulas for the means and variances of more general functions of random variables. In each case, the formulas involve the means, variances, and covariances of the individual variables in the function. Usually, these quantities are unknown, and they are estimated from sample or experimental data. Section 3.6 addresses the estimation problem.

2.7.1 Linear Combinations of Random Variables

Assume that the random variables X_1, X_2, \dots, X_k have means $\mu_1, \mu_2, \dots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ and that the covariance between X_i and X_j is σ_{ij} . Define the random variable Y to be a linear combination of X_1, X_2, \dots, X_k of the form

$$Y = a_0 + \sum_{i=1}^k a_i X_i \quad (2.121)$$

where a_0, a_1, \dots, a_k are arbitrary constants. The mean of Y is given by

$$\mu_Y = E[Y] = a_0 + \sum_{i=1}^k a_i \mu_i \quad (2.122)$$

and the variance of Y is given by

$$\sigma_Y^2 = \text{Var}[Y] = \sum_{i=1}^k a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j>i} a_i a_j \sigma_{ij} \quad (2.123)$$

When the X_i are from a k -variate normal distribution, then Y is normally distributed with mean and variance given by Equations 2.122 and 2.123. When the X_i are independent normally distributed random variables, then Y is normally distributed with mean and variance given by Equations 2.122 and 2.123, where all covariances σ_{ij} are zero.

Suppose, however, that the exact distributions of the individual random variables X_i are unknown but that they are independently distributed. In such cases, it is often possible to apply the Central Limit Theorem (Theorem 2.7). To do so requires that the number of individual terms $a_i X_i$ be sufficiently large to satisfy the rules of thumb given in Section 2.5.2.1 for applying the Central Limit Theorem. Note that $a_i X_i$ is a random variable with mean $a_i \mu_i$ and variance $a_i^2 \sigma_i^2$. When applicable, the Central Limit Theorem states that the distribution of Y can be approximated by a normal distribution with mean and variance given by Equations 2.122 and 2.123, where all covariances σ_{ij} are zero. In such cases, statistical procedures based on the normal distribution can be applied to make inferences about Y .

Example 2.10 In this example, given by Jaech (1973), the total weight of a fuel rod is the sum of the cladding, end-plug, spring, and fuel weights. Assume that these weights are random variables with means and variances given below.

Component weight	Mean weight, g	Variance, g ²
Cladding: X_1	$\mu_1 = 554.89$	$\sigma_1^2 = 16.8100$
Plug: X_2	$\mu_2 = 10.64$	$\sigma_2^2 = 0.0016$
Spring: X_3	$\mu_3 = 16.07$ g	$\sigma_3^2 = 0.0036$
Fuel: X_4	$\mu_4 = 1814.85$ g	$\sigma_4^2 = 6.0025$

Assuming that all covariances are zero, use Equations 2.122 and 2.123 to determine the mean and variance of the total weight of a fuel rod.

The total weight of a fuel rod is

$$Y = \sum_{i=1}^4 X_i$$

which is the form of Equation 2.121 with $a_0 = 0$ and all other $a_i = 1$. Thus, Equations 2.122 and 2.123 give

$$\mu_Y = 554.89 + 10.64 + 16.07 + 1814.85 = 2396.45 \text{ g}$$

and

$$\sigma_Y^2 = 16.8100 + 0.0016 + 0.0036 + 6.0025 = 22.8177 \text{ g}^2$$

The standard derivation of Y is

$$\sigma_Y = \sqrt{22.8177 \text{ g}^2} = 4.78 \text{ g}$$

which is in the same units as μ_Y .

2.7.2 General Functions of Random Variables

Assume that the random variables X_1, X_2, \dots, X_k have means $\mu_1, \mu_2, \dots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ and that the covariance between X_i and X_j is σ_{ij} . Define the random variable Y to be a known mathematical function of X_1, X_2, \dots, X_k , denoted by

$$Y = h(X_1, X_2, \dots, X_k) \quad (2.124)$$

When the pdf's of the individual X_i are known, it is possible to derive the pdf of Y using techniques of mathematical statistics. This pdf can then be used to derive exact expressions for the mean and variance of Y. Depending on the complexity of Y, however, the resulting mathematical expressions can be cumbersome to work with. When the pdf's of the individual X_i are unknown, it is not possible to derive the pdf of Y, and when Y is

other than a linear combination of the X_i , this generally precludes deriving exact expressions for the mean and variance of Y .

One approach, which provides approximate results and has been found useful in many practical applications, involves a first-order Taylor series expansion of Y about the point where each $X_i = \mu_i$. (The Taylor series expansion is presented in most basic calculus texts.) This provides an approximation of the function which is a linear combination of the X_i , and the techniques described in Section 2.7.1 can be applied to obtain approximate expressions for the mean and variance of Y .

The first-order Taylor series expansion of Y about the point where each $X_i = \mu_i$ is given by

$$Y \cong h(\mu_1, \mu_2, \dots, \mu_k) + \sum_{i=1}^k \left(\frac{\partial h}{\partial X_i} \right) (X_i - \mu_i) \quad (2.125)$$

where the partial derivatives $(\partial h / \partial X_i)$ are evaluated at $X_i = \mu_i$ and, thus, are constants. Because the μ_i are constants,

$$E[X_i - \mu_i] = E[X_i] - \mu_i = \mu_i - \mu_i = 0$$

and

$$\text{Var}[X_i - \mu_i] = \text{Var}[X_i] = \sigma_i^2$$

Also,

$$\text{Cov}[(X_i - \mu_i), (X_j - \mu_j)] = \text{Cov}(X_i, X_j) = \sigma_{ij}$$

Using these results together with Equations 2.122 and 2.123, approximate expressions for the mean and variance of Y are given by

$$\mu_Y \cong h(\mu_1, \mu_2, \dots, \mu_k) \quad (2.126)$$

and

$$\sigma_Y^2 \cong \sum_{i=1}^k \left(\frac{\partial h}{\partial X_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j>i} \left(\frac{\partial h}{\partial X_i} \right) \left(\frac{\partial h}{\partial X_j} \right) \sigma_{ij} \quad (2.127)$$

where the partial derivatives are evaluated at $X_i = \mu_i$.

The validity of these expressions depends upon how well the first-order Taylor series approximation represents Y . This can be investigated over the relevant range of the X_i variables by substituting values of the X_i into both Equations 2.124 and 2.125 and then comparing the resulting values of Y . Another problem is that the μ_i , σ_i^2 , and σ_{ij} are usually unknown and must be estimated from sample data. Substituting sample estimates for these unknown quantities in Equations 2.126 and 2.127 can yield estimates of μ_Y and σ_Y^2 with undesirable statistical properties. This is discussed further in Section 3.6. Thus, it is possible to obtain poor estimates of inadequate approximations. Generally, however, if the Taylor series approximation of Y is deemed adequate over the relevant range of the X_i variables, and if each σ_i is small relative to μ_i , the resulting estimates of μ_Y and σ_Y^2 are acceptable for practical use.

Example 2.11 In this example, presented by Jaech (1973), the amount of ^{235}U in a container of UO_2 powder is measured by determining the net weight of the powder and then drawing a sample of powder for analysis of percent uranium and percent ^{235}U . The random variables are:

X_1 = net weight of UO_2 powder, with mean $\mu_1 = 20.0$ kg and variance $\sigma_1^2 = (0.05)^2 \text{ kg}^2$.

X_2 = ratio of uranium to UO_2 (proportion of uranium) with mean $\mu_2 = 0.876$ and variance $\sigma_2^2 = (0.001)^2$.

X_3 = ratio of ^{235}U to uranium, with mean $\mu_3 = 0.0425$ and variance $\sigma_3^2 = (0.0002)^2$.

It is assumed that X_1 , X_2 and X_3 are independent, so that all covariances σ_{ij} are zero.

The estimated amount of ^{235}U in a container is given by

$$Y = X_1 X_2 X_3$$

Determine the mean and variance of Y .

The mean of Y is approximated by substituting μ_1 , μ_2 , and μ_3 into Equation 2.126 which gives

$$\mu_Y \cong (20.0)(0.876)(0.0425) = 0.7446 \text{ kg}$$

The variance of Y is approximated by applying Equation 2.127, where the partial derivatives, evaluated at $X_i = \mu_i$, are

$$\frac{\partial Y}{\partial X_1} = X_2 X_3 = (0.876)(0.0425) = 0.03723$$

$$\frac{\partial Y}{\partial X_2} = X_1 X_3 = (20.0)(0.0425) = 0.850$$

and

$$\frac{\partial Y}{\partial X_3} = X_1 X_2 = (20.0)(0.876) = 17.52$$

Thus, the variance of Y is approximately given by

$$\begin{aligned}\sigma_Y^2 &\cong (0.03723)^2 (0.05)^2 + (0.850)^2 (0.001)^2 \\ &\quad + (17.52)^2 (0.0002)^2 \\ &= 16.466 \times 10^{-6} \text{ kg}^2\end{aligned}$$

The standard deviation of Y is approximated by

$$\sigma_Y \cong \sqrt{16.466 \times 10^{-6} \text{ kg}^2} = 0.0041 \text{ kg}$$

REFERENCES

- American Heritage Dictionary Second College Edition*, 1982, Houghton Mifflin Company, Boston.
- Anderson, T. W., 1984, *Introduction to Multivariate Statistical Analysis*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Cramér, H., 1946, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey.
- Feller, W., 1968, *An Introduction to Probability Theory and Its Applications, Volume 1*, John Wiley & Sons, Inc., New York.
- Hahn, G. J., and S. S. Shapiro, 1967, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York.
- Hald, A., 1952, *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York.
- Hines, W. W., and D. C. Montgomery, 1980, *Probability and Statistics in Engineering and Management Science*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Hogg, R. V., and A. T. Craig, 1978, *Introduction to Mathematical Statistics*, 4th Ed., The Macmillan Company, New York.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Johnson, N. L., and S. Kotz, 1969, *Distributions in Statistics: Discrete Distributions*, John Wiley & Sons, Inc., New York.

- , and S. Kotz, 1970, *Distributions in Statistics: Continuous Univariate Distributions, Volumes 1 and 2*, John Wiley & Sons, Inc., New York.
- Kshirsagar, A. M., 1972, *Multivariate Analysis*, Marcel Dekker, Inc., New York.
- Larson, H. J., 1969, *Introduction to Probability Theory and Statistical Inference*, John Wiley & Sons, Inc., New York.
- Rohatgi, V. K., 1984, *Statistical Inference*, John Wiley & Sons, Inc., New York.
- Webster's Ninth New Collegiate Dictionary*, 1983, Merriam-Webster, Inc., Springfield, Massachusetts.

CHAPTER 3

Estimation

3.0 INTRODUCTION

In Chapter 2, the concepts of random variables, probability density functions (pdf), and cumulative distribution functions (cdf) are introduced. A random variable is a function which assigns a numerical value to each possible outcome of an experiment. The set of all possible numerical values or labels of a random variable is referred to as the **population** of values or labels. Information about the probability of observing the values or labels of a random variable is obtained from its pdf (or cdf). The pdf of a random variable is a function with one or more **parameters**. For example, the mean (μ) and the variance (σ^2) are the parameters that describe a normal pdf, whereas the probability of success (p) and the number of trials (n) are the parameters that describe a binomial distribution. Because the parameters of a pdf are usually unknown, it is necessary to make **inferences** about them based upon a **sample** of observed values of the random variable taken from the population. The concept of a sample is that only a portion of all possible observations is obtained from the population. An **inference** on a parameter involves reaching a conclusion or decision about a parameter based upon the information in a sample. There are two major types of **statistical inference: estimation and hypothesis testing**. The topic of estimation is introduced in this chapter; hypothesis testing is introduced in Chapter 4.

A **statistic** is a function of a sample; that is, it is an expression or formula that is used to calculate a numerical value from sample observations. An **estimator** of an unknown parameter is a statistic that yields a numerical estimate of the parameter. An **estimate** is the numerical value of an estimator calculated from a set of sample data. Because an estimator is a known mathematical function of random variables, it is a random variable. An estimate is an observed value of this random variable. Thus, an estimator itself has a probability distribution referred to as its **sampling distribution**.

In this chapter, two types of estimators are considered. A **point estimator** is a statistic that, when evaluated for a given set of sample data, results in a single numerical value for the estimate of an unknown parameter. An **interval estimator** is a statistic that, when evaluated for a given set of sample data, results in an interval with a specified probability of including the

true value of an unknown parameter. These random intervals are called **confidence intervals**. Often, a confidence interval for an unknown parameter can be developed from a point estimator of the parameter and the sampling distribution of this point estimator.

Empirical density functions and cdf's are discussed in Section 3.1. Some properties of estimators are discussed briefly in Section 3.3. Point and interval estimators of several population parameters introduced in Chapter 2 are discussed in Sections 3.4 and 3.5, respectively. Estimators for functions of random variables are considered in Section 3.6. Statistical tolerance limits for normal distributions are discussed in Section 3.7.

3.1 EMPIRICAL DISTRIBUTIONS

An approximation of a probability density function (pdf) developed from sample observations is known as an **empirical density function**. A similar approximation of a cumulative distribution function (cdf) is known as an **empirical cumulative distribution function**. These are discussed in Section 3.1.1.

Moments of a pdf are introduced in Section 2.3 and are referred to as **population moments**. Moments of an empirical density function are referred to as **sample moments** and are discussed in Section 3.1.2.

3.1.1 Empirical Density and Cumulative Distribution Functions

In situations where the functional form of a population pdf (or cdf) is unknown, sample observations are used to construct an approximation and to aid in determining the form of the pdf (or cdf).

Let X represent either a discrete or a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Let x_1, x_2, \dots, x_n represent the observed values of a random sample of size n taken from $f(x)$. Assume that the random sample is ordered from smallest to largest, and denote the ordered values by $x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(n)}$. In this ordered form, $x^{(i)}$, $i = 1, 2, \dots, n$, is referred to as the i^{th} **order statistic**. Then, the **empirical density function** of X is given by

$$g(x) = \begin{cases} \frac{1}{n} & \text{for } x = x^{(i)}, i = 2, \dots, n \\ 0 & \text{elsewhere} \end{cases} \quad (3.1)$$

The **empirical cumulative distribution function** of X is given by

$$G(x) = \begin{cases} 0 & \text{for } x < x_1 \\ \frac{k}{n} & \text{for } x^{(k)} \leq x < x^{(k+1)}, k = 1, 2, \dots, n-1 \\ 1 & \text{for } x \geq x_n \end{cases} \quad (3.2)$$

In words, the empirical cumulative distribution function $G(x)$ is equal to the fraction of the x_i 's that are less than or equal to x for each x , $-\infty < x < \infty$. Illustrations of the empirical cumulative distribution function are found in Examples 9.17 and 9.18 of Chapter 9.

Empirical density and cumulative distribution functions provide simple approximations to unknown pdf's and cdf's. A simple approximation to pdf's, the histogram, is considered in Chapter 1. A histogram often provides a sufficient approximation for graphical display and exploratory purposes. For more complicated applications where a functional form is required and none of the well-known distributions are appropriate, **Johnson** or **Pearson approximations** are often used. Hahn and Shapiro (1967) discuss techniques for determining which well-known distribution might adequately represent a set of sample observations. They also discuss the development of Johnson and Pearson approximations when no well-known distribution is appropriate. The goodness-of-fit tests discussed in Section 9.6 are also applicable for this purpose.

3.1.2 Sample Moments

The k^{th} population moment of the random variable X is defined in Section 2.3 as

$$\mu_k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx \quad (3.3)$$

for X continuous and as

$$\mu_k = E[X^k] = \sum_x x^k f(x) \quad (3.4)$$

for X discrete. In both cases $f(x)$ is the pdf of X . In an analogous manner, the k^{th} **sample moment** is defined as

$$m_k = \sum_{i=1}^n x_i^k g(x_i) = \sum_{i=1}^n \frac{x_i^k}{n} \quad (3.5)$$

where the empirical density function $g(x)$ is given by Equation 3.1. Note that this equation provides sample moments for both discrete and continuous random variables X . The first sample moment ($k = 1$) is referred to as the **sample mean** and is denoted by \bar{x} .

The k^{th} moment about the population mean is defined in Section 2.3 as $\mu'_k = E[(X - \mu)^k]$. The k^{th} sample moment of X about the sample mean is given by

$$m'_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^k}{n} \quad (3.6)$$

3.2 PROPERTIES OF POINT ESTIMATORS

There are a number of properties which are used to judge estimators of population parameters. Several of these are discussed below.

1. A point estimator $\hat{\gamma}$ of the parameter γ is said to be **unbiased** if $E[\hat{\gamma}] = \gamma$; that is, if the mean of the sampling distribution of the estimator $\hat{\gamma}$ equals the population parameter γ . When $E[\hat{\gamma}] \neq \gamma$, $\hat{\gamma}$ is said to be a **biased** estimator of γ .

2. A point estimator $\hat{\gamma}$ is said to have **minimum variance** with respect to a class of estimators if the estimator has a variance no larger than the variance of any other estimator of the parameter γ in the class. For example, the estimator in the class of unbiased estimators with minimum variance is the **minimum variance unbiased estimator**.

3. Many estimators are a function of the sample size n . An estimator $\hat{\gamma}$ is said to be **consistent** if it becomes arbitrarily close to γ as n increases and approaches infinity. Mathematically, this is written as

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\gamma} - \gamma| \geq \epsilon) = 0 \quad \text{for any } \epsilon > 0$$

An estimator that improves with increasing sample size (a consistent estimator) is both intuitively and theoretically appealing.

4. A point estimator is said to be **robust** if it is not seriously affected by departures from distributional assumptions or by potential extreme sample observations.

The above properties of estimators are referred to in Section 3.4 during the presentation of point estimators for several common population parameters of interest. Although it is desirable to develop estimators having all of the above properties, this is not always possible. In a specific situation, there may be reasons why an estimator missing one or more of the previously mentioned properties is preferred. The trade-offs among various properties may determine which estimator should be used.

A common measure of performance of a point estimator $\hat{\gamma}$ is the **mean-squared error**, defined by

$$\text{MSE}(\hat{\gamma}) = E[(\hat{\gamma} - \gamma)^2] \quad (3.7)$$

which can be written as

$$\begin{aligned} E[(\hat{\gamma} - \gamma)^2] &= E[(\hat{\gamma} - E[\hat{\gamma}])^2] + (E[\hat{\gamma}] - \gamma)^2 \\ &= \text{Var}(\hat{\gamma}) + (\text{Bias}[\hat{\gamma}])^2 \end{aligned} \quad (3.8)$$

If $\hat{\gamma}$ is an unbiased estimator of γ , $\text{Bias}[\hat{\gamma}] = 0$ and $\text{MSE}(\hat{\gamma}) = \text{Var}(\hat{\gamma})$.

Mean-squared error is sometimes used to compare two or more estimators, with preference given to those estimators with smaller MSE's.

3.3 METHODS FOR DERIVING ESTIMATORS

In this section, four commonly used methods for deriving point estimators are discussed. These are the **method of moments**, **least squares**, **maximum likelihood**, and **Bayesian estimation**. Depending on the distribution being sampled, these methods may yield identical or different estimators.

3.3.1 Method of Moments

The method of moments devised by Karl Pearson in 1894 is one of the oldest methods of deriving estimators.

Assume that k parameters of a population distribution are to be estimated. The method consists of expressing the first k population moments (Section 2.3, Section 3.1.2, and Equations 3.3 and 3.4) in terms of these k parameters, equating them to the corresponding sample moments, and then solving the resulting equations for the k parameters. This gives k formulas which express each of the k parameters as a function of the sample moments. These formulas are the method of moments estimators of the k parameters. This method is illustrated by the following example.

Example 3.1 An audit of a manual record-keeping system is to be performed. For a random sample of n records, each record is examined to determine whether it does or does not contain one or more errors. Develop an estimator of the probability of a defective record, i.e., a record containing one or more errors.

The result of examining a single record is a random variable X , where $X = 1$ if the record contains one or more errors and $X = 0$ if the record contains no errors. Assume that $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$ for all records in the population. Note that X is a special case of the binomial distribution (Section 2.4.2) with $n = 1$. The parameter (probability) p is estimated by the method of moments by equating the first population moment to the first sample moment. Because $E[X_i] = p$ is the first population moment, equating $E[X_i]$ to the first sample moment gives

$$p = m_1 = \sum_{i=1}^n x_i/n = \frac{1 + 0 + 1 + \dots + 0 + 1}{n} = \frac{y}{n}$$

where

$$y = \sum_{i=1}^n x_i$$

is the number of defective records observed in the random sample. Note that y is an observation on a binomially distributed random variable Y with mean np and variance $np(1 - p)$. Hence, $\hat{p} = Y/n$ is an unbiased estimator of p , because

$$E[\hat{p}] = E\left[\frac{Y}{n}\right] = \frac{1}{n} E[Y] = \frac{1}{n} (np) = p$$

3.3.2 Least Squares

Least squares is a specialized estimation technique that is widely used for fitting equations to data; it is further discussed in Chapter 6.

Consider the sum of squares

$$\sum_{i=1}^n (x_i - E[X_i])^2 \quad (3.9)$$

where x_i is an observation on the random variable X_i . The expectation of X_i is $E[X_i]$, which depends upon the unknown parameters to be estimated. Least squares estimation attempts to minimize the sum of squares in Equation 3.9 over possible values of the parameters. If a unique minimum exists, the corresponding values of the parameters are their **least squares estimates**. Note that for a unique minimum to exist, the number of observations n in the sample must be equal to or larger than the number of parameters being estimated.

In a random sample where X_1, X_2, \dots, X_n are not independent and identically distributed, the sum of squares in Equation 3.9 must be modified before minimization. This generalized least squares problem is discussed in Chapter 6, where the method of least squares is used extensively in estimating regression equation parameters.

3.3.3 Maximum Likelihood

Consider a random sample X_1, X_2, \dots, X_n from a pdf $f(x, \theta)$ that depends upon an unknown parameter θ . The joint pdf of X_1, X_2, \dots, X_n (Section 2.6), also a function of θ , is referred to as the **likelihood function** and is given by

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta) \quad (3.10)$$

Using calculus techniques, it may be possible to find a value $\hat{\theta}$ that maximizes the likelihood function when substituted for θ . If so, then $\hat{\theta}$ is the **maximum likelihood estimate** of the parameter θ . This technique is illustrated in the following example.

Example 3.2 Derive the maximum likelihood estimator of p for the problem described in Example 3.1.

The pdf of X is

$$f(x) = p^x (1 - p)^{1-x} \quad x = 0, 1$$

The likelihood function of a random sample X_1, X_2, \dots, X_n is

$$\begin{aligned} L(p; x_1, x_2, \dots, x_n) &= p^{x_1} (1 - p)^{1-x_1} \dots p^{x_n} (1 - p)^{1-x_n} \\ &= p^y (1 - p)^{n-y} \end{aligned}$$

where

$$y = \sum_{i=1}^n x_i$$

The value of p which maximizes the likelihood function is found by differentiating the logarithm† of $L(p; x_1, x_2, \dots, x_n)$ with respect to p , setting the derivative equal to zero, and then solving for p . These steps are illustrated below.

$$\ln L(p; x_1, x_2, \dots, x_n) = y (\ln p) + (n - y) \ln(1 - p)$$

$$\frac{d \ln L}{dp} = \frac{y}{p} - \frac{n - y}{1 - p} = 0$$

$$y(1 - p) - p(n - y) = 0$$

† Maximizing the logarithm of a function is equivalent to maximizing the function itself. It is sometimes easier to maximize the logarithm of the function.

$$np = y$$

$$p = y/n$$

Hence, $\hat{p} = Y/n$ is the maximum likelihood estimator of p . Note that the same estimator was obtained in Example 3.1 by using the method of moments.

Maximum likelihood estimation is easily extended to estimating k parameters $\theta_1, \theta_2, \dots, \theta_k$ based upon a random sample X_1, \dots, X_n from a pdf $f(x; \theta_1, \theta_2, \dots, \theta_k)$. The likelihood function is still obtained as the joint pdf of X_1, \dots, X_n but is now considered as a function of the parameters $\theta_1, \theta_2, \dots, \theta_k$. The maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ of the parameters are generally those values of $\theta_1, \theta_2, \dots, \theta_k$ that solve the k equations

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

or equivalently

$$\frac{\partial (\ln L)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

In some applications, the possible values of the parameters $\theta_1, \theta_2, \dots, \theta_k$ are constrained. For example, suppose that a negative value of θ_1 is meaningless. In such cases, the likelihood function is maximized subject to specified constraints on the values of the parameters. The resulting estimates are usually called **constrained maximum likelihood estimates**.

3.3.4 Bayesian Estimation

In some estimation problems, prior knowledge about one or more of the population parameters may exist. For example, in the records-auditing illustration of Examples 3.1 and 3.2, prior work in investigating error frequencies might suggest that a smaller value of p is more likely than a larger value. **Bayesian estimation** uses such prior information in addition to sample observations for estimating unknown parameters. Prior information about the parameters to be estimated is incorporated by assuming that p has a specified distribution (referred to as a **prior distribution**).

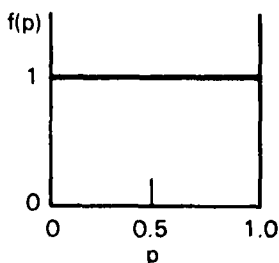
The description of the actual derivation process for obtaining Bayesian estimators is beyond the scope of this brief discussion. The interested reader can refer to DeGroot (1970); Martz and Waller (1982); Larson (1969); Lindgren (1976); Hogg and Craig (1970); Hoel, Port, and Stone (1970); or other texts on mathematical statistics.

Example 3.3 Consider the following estimators of p for the problem described in Examples 3.1 and 3.2.

Prior Distribution of p

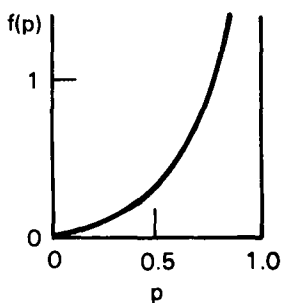
Bayesian Estimator, \hat{p}

(1) Uniform



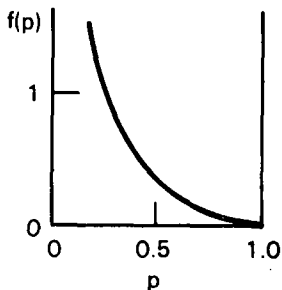
$$\hat{p}_1 = \frac{Y + 1}{n + 2}$$

(2) Beta ($\alpha=3, \beta=1/2$)



$$\hat{p}_2 = \frac{Y + 3}{n + 3.5}$$

(3) Beta ($\alpha=1/2, \beta=3$)



$$\hat{p}_3 = \frac{Y + 1/2}{n + 3.5}$$

Note that $\hat{p}_3 < \hat{p}_1 < \hat{p}_2$, as might be expected from viewing the graphs of the prior distributions. Also, none of the three estimators is unbiased.

3.4 POINT ESTIMATORS

Point estimators for several of the population parameters introduced in Chapter 2 are presented in this section. These estimators, for the most part, are obtained using one of the methods for deriving estimators discussed in Section 3.3. Because only the estimators themselves are of interest here, their derivations are not presented. Examples illustrating the use of estimators presented in this section are deferred until Section 3.5. Examples of interval estimators in that section also illustrate the use of point estimators.

3.4.1 Estimators for the Mean and Variance of a Normal Distribution

Consider a normally distributed random variable X with unknown population mean μ and variance σ^2 . Let X_1, X_2, \dots, X_n represent a random sample of size n . The method of moments, least squares, and maximum likelihood estimators of μ are identical and are given by the **sample mean**

$$\bar{X} = \sum_{i=1}^n X_i/n \quad (3.11)$$

This is the minimum variance unbiased estimator for μ and is also consistent.

For any distribution with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, \bar{X} is an unbiased and consistent estimator for μ . The variance of \bar{X} is σ^2/n . The standard deviation of \bar{X} is often referred to as its standard error and is equal to σ/\sqrt{n} .

In some situations, estimators of μ other than \bar{X} may be preferred. For example, \bar{X} is sensitive to extreme observations in the sample; i.e., it is not **robust**. The sample median, introduced in Section 1.2.1.3, is a more robust estimator of μ than \bar{X} . For the normal distribution, the sample median is an unbiased estimator of μ but always has larger variance than \bar{X} for $n > 2$. Other robust estimators of μ are presented and discussed by Andrews et al. (1972).

We now consider point estimators for the unknown population variance σ^2 based on a random sample X_1, X_2, \dots, X_n . The method of moments† and maximum likelihood estimators are identical and are given by

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2/n}{n} \quad (3.12)$$

† Even if the random variable X is not normally distributed, Equation 3.12 supplies the method of moments estimator of σ^2 .

Note that $\hat{\sigma}^2$ is biased because it can be shown that

$$E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

An unbiased estimator of σ^2 is given by

$$S^2 = \frac{n}{n-1} \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2/n}{n-1} \quad (3.13)$$

Both $\hat{\sigma}^2$ and S^2 are consistent estimators for σ^2 .

For any distribution with $\text{Var}(X) = \sigma^2$, it can be shown that $E[S^2] = \sigma^2$, i.e., S^2 is an unbiased estimator of σ^2 as stated previously. When the X_i are a random sample from a normal distribution with variance σ^2 , the variance of S^2 is given by

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1} \quad (3.14)$$

Equations 3.12 and 3.13 assume that the random sample X_1, X_2, \dots, X_n is drawn either from a finite population with replacement or from an infinite population. Chapter 8 provides formulas and techniques for cases where the random sample is drawn without replacement from a finite population.

The estimators $\hat{\sigma}^2$ and S^2 are not particularly robust because they are inflated by extreme sample observations. Although more robust estimators exist, such as the interquartile range (Section 1.3.4), they are not often used. In practice, the influence of confirmed outlying observations on estimators of σ^2 can be negated by deleting the outlying observations from the sample before calculating the point estimate. Techniques for detecting outlying observations are discussed in Section 10.4.

3.4.2 Estimators for the Binomial Distribution Parameter

Point estimators for p , one of the parameters of a binomial distribution, are presented in this section. When sampling from a binomial distribution, there are only two possible outcomes on each "trial," often referred to as "success" or "failure." The parameter p is the true unknown probability of a success in the population. Using the convention that $X = 1$ for a success and $X = 0$ for a failure, the method of moments, least squares, and maximum likelihood estimators for p are identical and are given by

$$\hat{p} = \frac{Y}{n} = \frac{\sum_{i=1}^n X_i}{n} \quad (3.15)$$

where Y is the number of successes observed in the random sample of size n . The derivations of Equation 3.15, using the method of moments and maximum likelihood estimation, are given, respectively, in Sections 3.3.1 and 3.3.3. The estimator \hat{p} given by Equation 3.15 is an unbiased and consistent estimator of p .

If some prior information exists, biased Bayesian estimators of p such as those in Example 3.3 might be considered.

3.4.3 Estimators for Other Distributions

Maximum likelihood estimators for the parameters of the hypergeometric, Poisson, uniform, gamma, and lognormal distributions are summarized in Table 3.1. Readers interested in the properties of these estimators (unbiasedness, etc.) are referred to Johnson and Kotz (1969; 1970).

3.5 SAMPLING DISTRIBUTIONS AND INTERVAL ESTIMATORS

In Section 3.3, methods are presented for obtaining point estimators of an unknown population parameter. For each possible sample of a given size, an estimator may yield a different point estimate. All such possible estimates form a probability distribution referred to as the **sampling distribution** of the estimator. That is, the estimator is a random variable with its own distribution.

Starting with a point estimator, its sampling distribution is employed to derive an **interval estimator** of the unknown parameter.† An interval estimator is a random interval (an interval with at least one end-point being a random variable) with a prespecified probability that it will include the true value of the unknown parameter. The probability structure is provided by the sampling distribution of the estimator. These random intervals are usually called **confidence intervals**.

Consider the interval $[L, U]$, where either L or U or both are random variables. Then, if

$$\Pr(L \leq \theta \leq U) = 1 - \alpha \quad (3.16)$$

† Although all of the interval estimators discussed in this section are developed in this manner, there are other ways to develop interval estimators. See Rohatgi (1984) for a discussion of this topic.

TABLE 3.1

Maximum Likelihood Estimators for Parameters of Selected Distributions

Distribution	Parameter	Maximum likelihood estimator	Notes
Hypergeometric (Section 2.4.1)	D	$\hat{D} = \frac{X(N+1)}{n}$	D = number of defects in population \hat{D} is rounded down to next integer X = number of defects in sample N = population size n = sample size
Poisson (Section 2.4.3)	μ	$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n}$	μ = mean (and variance) of Poisson distribution X_i = i^{th} observation in the random sample n = sample size
Uniform (Section 2.5.1)	α β	$\hat{\alpha} = X^{(1)}$ $\hat{\beta} = X^{(n)}$	$X^{(1)}$ = the smallest value in the sample $X^{(n)}$ = the largest value in the sample
Gamma* (Section 2.5.4)	r	$\hat{r} = r^{-1} \left[\frac{1}{n} \sum_{i=1}^n \ln X_i \right]$	r = shape parameter λ = scale parameter X_i = i^{th} observation in the random sample n = sample size \bar{X} = sample mean
	λ	$\hat{\lambda} = \frac{\hat{r}}{\bar{X}}$	
	μ_Y	$\hat{\mu}_Y = \bar{Y}$	$Y = \ln X \sim N(\mu_Y, \sigma_Y^2)$ $Y_i = \ln X_i$
Lognormal (Section 2.5.3)	σ_Y^2	$\hat{\sigma}_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$	$\bar{Y} = \sum_{i=1}^n Y_i/n$ X_i = i^{th} observation in the random sample n = sample size

*See Johnson and Kotz (1970) for a discussion of the inverse function f^{-1} .

the interval $[L, U]$ is called a $100(1 - \alpha)\%$ **confidence interval for the parameter θ** . In practice, a random sample is taken from the population, and values ℓ and u are calculated. Although the interval $[\ell, u]$ is referred to as a $100(1 - \alpha)\%$ confidence interval, care must be taken in the interpretation of the “percent confidence” statement. For example, suppose $[\ell, u]$ is a 95% confidence interval for θ . This *does not* mean there is a 95% chance that θ is between ℓ and u . Given ℓ and u , θ is either between them (100% chance) or it is not between them (0% chance); we do not know. It *does* mean, that if repeated samples are collected and ℓ and u are computed for each sample, 95% of the estimated intervals $[\ell, u]$ can be expected to include θ . In practice, only one sample is collected, and it may or may not produce one of the 95 out of 100 intervals which does include θ .

In this section, interval estimators are considered for the mean and variance of the normal distribution and for the parameter p of a binomial distribution. Because it is often of interest to compare these parameters for two populations, interval estimators for differences of means, ratios of variances, and differences of proportions are also considered. The sampling distributions required to develop the interval estimators are also presented. The well-known F, chi-square, and Student's t-distributions are introduced in this manner. The sampling distributions presented in this section also play an important part in the development of hypothesis tests in Chapter 4 and subsequent chapters. For developmental purposes, interval estimation of variances (for normal distributions) is considered first.

3.5.1 Variances and Ratios of Variances for the Normal Distribution

The point estimator of σ^2 most commonly used in nuclear material accounting applications is S^2 given by Equation 3.13. Although the sampling distribution of the random variable S^2 is known, the sampling distribution of a function of S^2 is more tractable for developing an interval estimator for σ^2 . The function $(n - 1)S^2/\sigma^2$ has a chi-square distribution with $n - 1$ degrees of freedom. Notationally, this is written

$$\chi^2 = \frac{(n - 1)S^2}{\sigma^2} \sim \chi^2(n - 1) \quad (3.17)$$

The pdf of a chi-square random variable with ν degrees of freedom is a special case of the gamma pdf (Table 2.5 in Section 2.5.4) where $\lambda = \frac{1}{2}$ and $r = \nu/2$. The mean and variance of a chi-square distribution with ν degrees of freedom are $E[\chi^2] = \nu$ and $\text{Var}(\chi^2) = 2\nu$, respectively; see Graybill (1976) for a derivation. Quantiles $\chi_p^2(\nu)$ of the distribution for various values of ν and p are given in Table A4 of Appendix A. Use of Table A4 is illustrated in Example 3.4.

Based upon the sampling distribution of χ^2 given by Equation 3.17, a $100(1 - \alpha)\%$ two-sided confidence interval for σ^2 is developed. Although the chi-square distribution is not symmetric, it is common practice to assign probability $\alpha/2$ to each tail in developing the two-sided confidence interval. Thus, a $100(1 - \alpha)\%$ two-sided confidence interval is given by

$$\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \quad (3.18)$$

A $100(1 - \alpha)\%$ two-sided confidence interval for the population standard deviation is given by

$$\{[(n-1)S^2]/[\chi_{1-\alpha/2}^2(n-1)]\}^{1/2} \leq \sigma \leq \{[(n-1)S^2]/[\chi_{\alpha/2}^2(n-1)]\}^{1/2} \quad (3.19)$$

A one-sided $100(1 - \alpha)\%$ **upper confidence limit** for σ^2 or σ is obtained by replacing $\alpha/2$ with α in the right-hand side of Equation 3.18 or 3.19 and setting the left-hand side equal to zero. A $100(1 - \alpha)\%$ **lower confidence limit** for σ^2 or σ , although rarely used in practice, is obtained by replacing $\alpha/2$ with α in the left-hand side of Equation 3.18 or 3.19 and setting the right-hand side equal to ∞ .

Example 3.4 A facility wishes to verify that the variability in the weights of completed fuel rods is sufficiently low that no problems will arise with respect to the total weight of the rods in an assembled fuel element. A sample of 12 rods is weighed, with the following results (in grams).

2396.85, 2397.41, 2395.20, 2397.75, 2389.27, 2391.08,
2388.31, 2394.52, 2393.79, 2403.17, 2397.01, 2388.82

Estimate the variability in the fuel rod weights and determine a 95% one-sided upper confidence limit.

A point estimate of σ^2 , obtained by substituting the sample observations into Equation 3.13, is given by

$$\begin{aligned} s^2 &= \frac{(2396.85^2 + \dots + 2388.82^2) - (2397.86 + \dots + 2388.82)^2/12}{12 - 1} \\ &= \frac{68,799,852.66 - (28,733.18)^2/12}{11} = \frac{216.58}{11} = 19.69 \text{ g}^2 \end{aligned}$$

A point estimate of the population standard deviation σ is given by

$s = \sqrt{19.69 \text{ g}^2} = 4.44 \text{ g}$. A 95% one-sided upper confidence limit for σ is obtained by using the right-hand side of Equation 3.19, with the quantile $\chi_{\alpha/2}^2(n-1)$ replaced by $\chi_{\alpha}^2(n-1)$. Because $\alpha = 0.05$ and $n = 12$, we have $\chi_{\alpha}^2(n-1) = \chi_{0.05}^2(11)$. Entering Table A4 with $p = 0.05$ and $\nu = 11$, the value $\chi_{0.05}^2(11) = 4.57$ is obtained. By substituting the above quantities into the right-hand side of Equation 3.19, the 95% upper confidence limit for σ is computed as

$$\sigma \leq \{[(12-1)s^2]/[\chi_{0.05}^2(12-1)]\}^{1/2}$$

$$\sigma \leq \{[11(19.69)]/4.57\}^{1/2}$$

$$\sigma \leq 6.88 \text{ g}$$

Thus, we are 95% confident that the standard deviation of fuel rod weights in this population does not exceed 6.88 g.

Now suppose that the variances of two populations are to be considered. Let random variables X_1 and X_2 have population variances σ_1^2 and σ_2^2 , respectively. If random samples of sizes n_1 and n_2 are selected from X_1 and X_2 , unbiased point estimators of the population variances are individually given by Equation 3.13 and are denoted by S_1^2 and S_2^2 . Consider the random variable

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{[(n_1-1)S_1^2/\sigma_1^2]/(n_1-1)}{[(n_2-1)S_2^2/\sigma_2^2]/(n_2-1)} = \frac{U/(n_1-1)}{V/(n_2-1)} \quad (3.20)$$

where $U \sim \chi^2(n_1-1)$ and $V \sim \chi^2(n_2-1)$. The random variable F is said to have an **F-distribution with (n_1-1) and (n_2-1) degrees of freedom**. This is denoted by $F \sim F(n_1-1, n_2-1)$. The functional form of the F-distribution is not required for the applications presented in this text and is not given here. The interested reader is referred to Graybill (1976) for a detailed presentation. Quantiles $F_p(\nu_1, \nu_2)$ are given in Table A6 of Appendix A for various values of p , ν_1 , and ν_2 . Note that $F_{1-p}(\nu_1, \nu_2) = 1/F_p(\nu_2, \nu_1)$. An illustration of the use of Table A6 is given in Example 3.5.

Based upon the sampling distribution of the random variable F in Equation 3.20, a $100(1-\alpha)\%$ confidence interval for the ratio σ_1^2/σ_2^2 of the population variances is given by

$$\frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)} \frac{S_1^2}{S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{\alpha/2}(n_1-1, n_2-1)} \frac{S_1^2}{S_2^2} \quad (3.21)$$

A $100(1 - \alpha)\%$ two-sided confidence interval for the ratio of the population standard deviations is given by

$$\begin{aligned} & \{[1/F_{1-\alpha/2}(n_1-1, n_2-1)] (S_1^2/S_2^2)\}^{1/2} \leq \frac{\sigma_1}{\sigma_2} \\ & \leq \{[1/F_{\alpha/2}(n_1-1, n_2-1)] (S_1^2/S_2^2)\}^{1/2} \end{aligned} \quad (3.22)$$

A one-sided $100(1 - \alpha)\%$ confidence limit for σ_1^2/σ_2^2 or σ_1/σ_2 is obtained from Equation 3.21 or 3.22 by replacing $\alpha/2$ with α and choosing the appropriate side of the two-sided interval, depending upon whether an upper or lower confidence limit is to be computed.

Example 3.5 Several containers from each of two batches of UO_2 powder are measured for percent uranium with the following results.

Batch 1	Batch 2
87.659	87.692
87.626	87.614
87.637	87.662
87.662	87.650
87.619	87.675
87.630	87.687
87.640	87.661
87.621	87.691
	87.676
	87.664

Estimate the ratio of the population standard deviations for the two batches, and construct a 90% confidence interval for the ratio.

A point estimate of σ_1/σ_2 is given by $s_1/s_2 = 0.016/0.023 = 0.696$, where Equation 3.13 is used to compute s_1 and s_2 . A confidence interval for σ_1/σ_2 is computed from Equation 3.22. For $n_1 = 8$, $n_2 = 10$, and $\alpha = 0.10$, the quantiles required are

$$F_{1-\alpha/2}(n_1-1, n_2-1) = F_{0.95}(7, 9)$$

and

$$F_{\alpha/2}(n_1-1, n_2-1) = F_{0.05}(7, 9)$$

For $F_{0.95}(7, 9)$, we have $p = 0.95$, $\nu_1 = 7$, and $\nu_2 = 9$ in the notation of Table A6. The value $F_{0.95}(7, 9) = 3.29$ is obtained by locating the correct ν_1 across the top of the table, the correct ν_2 on the left margin, and the correct p value within the (ν_1, ν_2) combination. Because Table A6 does not include quantiles $F_p(\nu_1, \nu_2)$ for $p < 0.50$, the quantile $F_{0.05}(7, 9)$ is obtained from the relationship

$$F_{0.05}(7, 9) = \frac{1}{F_{0.95}(9, 7)} = \frac{1}{3.68} = 0.27$$

By substituting the above quantities into Equation 3.22, the 90% confidence interval is computed as

$$\{(1/3.29)[(0.016)^2/(0.023)^2]\}^{1/2} \leq \frac{\sigma_1}{\sigma_2} \leq \{(1/0.27)[(0.016)^2/(0.023)^2]\}^{1/2}$$

$$0.384 \leq \frac{\sigma_1}{\sigma_2} \leq 1.339$$

That is, a 90% confidence interval for the ratio σ_1/σ_2 is (0.384, 1.339).

3.5.2 Means and Differences of Means for the Normal Distribution

In Section 3.4.1, the sample mean \bar{X} is presented as the most frequently used point estimator of the population mean μ . We now consider the sampling distribution of \bar{X} and interval estimators developed from it.

If the random variable X is distributed as $N(\mu, \sigma^2)$, then by using the results of Section 2.7.1, the sample mean \bar{X} is distributed as $N(\mu, \sigma^2/n)$. If the random variable X is not normally distributed, then, according to the Central Limit Theorem presented in Section 2.5.2.1, \bar{X} is approximately distributed as $N(\mu, \sigma^2/n)$ if the sample size n is sufficiently large. Some rules of thumb are given in that section for deciding when n is large enough for this normal approximation to yield valid results.

With $\bar{X} \sim N(\mu, \sigma^2/n)$, it follows that

$$Z = (\bar{X} - \mu)/(\sigma^2/n)^{1/2} \sim N(0, 1) \quad (3.23)$$

as discussed in Section 2.5.2. Based upon this sampling distribution (assuming that σ^2 is known),† a $100(1 - \alpha)\%$ confidence interval for the unknown population mean μ is given by

$$\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (3.24)$$

where $z_{1-\alpha/2}$ is obtained from Table A3. A specific quantile z_p is found by locating the desired p value in the body of Table A3 and then reading the quantile value from the margins. To illustrate, suppose that the value of $z_{0.975}$ must be obtained. First, 0.975 is located in the body of Table A3. Then $z_{0.975} = 1.96$ is read from the margins. The digits 1.9 come from the left margin, while the digit 6 (in the hundreds place) comes from the top margin. If the exact desired value of p is not in the body of the table, the closest value may be chosen, or an interpolation procedure may be used.

If σ^2 is not known, it can be replaced in Equation 3.23 with its point estimator S^2 which is given by Equation 3.13. Then,

$$T = \frac{\bar{X} - \mu}{\sqrt{(S^2/n)}} = \frac{(\bar{X} - \mu)/\sqrt{\sigma^2/n}}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \frac{Z}{\sqrt{V/(n-1)}} \quad (3.25)$$

where $Z \sim N(0, 1)$ from Equation 3.23 and $V = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ from Equation 3.17. This new random variable T is said to have **Student's t-distribution with $(n-1)$ degrees of freedom**, which is denoted by $T \sim t(n-1)$. The form of the pdf of T is not required for the applications given in this book and is not given here. Quantiles $t_p(\nu)$ of the pdf are given in Table A5 in Appendix A for various values of p and ν .

Based upon the sampling distribution of T from Equation 3.25, a $100(1 - \alpha)\%$ confidence interval for the population mean μ is given by

$$\bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \quad (3.26)$$

A one-sided $100(1 - \alpha)\%$ confidence limit for μ is obtained from Equation 3.24 or 3.26 by substituting α for $\alpha/2$ and using the appropriate side of the two-sided confidence interval. Alternately, each side of a $100(1 - \alpha)\%$ confidence interval is a $100(1 - \alpha/2)\%$ one-sided confidence limit.

† Although a population parameter such as σ^2 is usually unknown, previous experience may yield an estimate which can often be treated as the true value.

Note that the confidence intervals given in Equations 3.24 and 3.26 are exact $100(1 - \alpha)\%$ confidence intervals when X is normally distributed. These confidence intervals, however, are only approximate $100(1 - \alpha)\%$ confidence intervals if X is not normally distributed.

Example 3.6 Calculate a 98% confidence interval for the population mean using the data from Example 3.4.

The desired confidence interval is given by Equation 3.26, assuming that the population variance is unknown. By entering Table A5 with $p = 1 - \alpha/2 = 0.99$ and $n = 12$, the value $t_{0.99}(11) = 2.72$ is obtained. The confidence interval is then computed as

$$2394.43 - 2.72 \frac{4.44}{\sqrt{12}} \leq \mu \leq 2394.43 + 2.72 \frac{4.44}{\sqrt{12}}$$

$$2394.43 - 3.49 \leq \mu \leq 2394.43 + 3.49$$

$$2390.94 \leq \mu \leq 2397.92$$

That is, a 98% confidence interval for μ is (2390.94 g, 2397.92 g).

To illustrate one-sided confidence limits, one can be 99% confident that $\mu \leq 2397.92$. Similarly, one can be 99% confident that $\mu \geq 2390.94$.

Another type of interval estimation problem involves calculating the sample size required to yield a $100(1 - \alpha)\%$ confidence interval with a specified width. For the confidence interval given by Equation 3.24, the half-width w is given by

$$w = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (3.27)$$

If σ is known (or a reliable estimate from previous data is available), then the sample size required to yield a $(1 - \alpha)\%$ confidence interval with half-width w is given by solving Equation 3.27 for n . This gives

$$n = \left(z_{1-\alpha/2} \frac{\sigma}{w} \right)^2 \quad (3.28)$$

Because the sample size is an integer, n is rounded to the nearest integer.

When it is desirable to be conservative, n is rounded up. Equation 3.28 can also be used to calculate the sample size required for a one-sided confidence limit by substituting $z_{1-\alpha}$ in place of $z_{1-\alpha/2}$.

For cases where σ^2 is not known and a reliable estimate does not exist prior to sampling, Kendall and Stuart (1973) present a two-stage method. Briefly, a random sample of specified size is taken, and the resulting data are used to estimate σ^2 . This estimate is used to determine whether additional sampling is necessary, and if so, how many additional observations are required. If required, the additional sampling is completed, and the new data are combined with the initial data to construct a confidence interval for μ . For specific details of the procedure, see Kendall and Stuart (1973).

Example 3.7 For a particular UO_2 manufacturing process, historical process data indicate that the standard deviation of percent uranium in the pellets is 0.025%. A 95% two-sided confidence interval with half-width 0.01% for the true mean percent uranium value for a lot of 10,000 pellets is desired. Determine how many pellets must be measured.

The desired sample size is given by Equation 3.28, where $\alpha = 0.05$, $\sigma = 0.00025$, and $w = 0.0001$; that is,

$$n = \left(z_{0.975} \frac{0.00025}{0.0001} \right)^2 = [1.96(2.50)]^2 = 24.01 \cong 24$$

A sample of 24 pellets should be measured.

In some nuclear material control applications, it is necessary to estimate the difference between two population means. Let X_1 and X_2 be two random variables with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and suppose that $\mu_1 - \mu_2$ is to be estimated. A point estimator for this difference is given by $\bar{X}_1 - \bar{X}_2$. A confidence interval for $\mu_1 - \mu_2$ can be based on this point estimator, but the precise form depends upon the relationship of σ_1^2 and σ_2^2 . One form results if $\sigma_1^2 = \sigma_2^2$, whereas another form results if $\sigma_1^2 \neq \sigma_2^2$. If σ_1^2 and σ_2^2 are unknown, a test based upon their sample estimates may be performed to determine whether or not $\sigma_1^2 = \sigma_2^2$. A hypothesis test for this purpose is discussed in Section 4.3.3.

Now, assume that σ_1^2 and σ_2^2 are unknown, but independent random samples of size n_1 and n_2 are available. If $\sigma_1^2 = \sigma_2^2$, a $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\begin{aligned}
 (\bar{X}_1 - \bar{X}_2) - t_{1-\alpha/2}(n_1 + n_2 - 2) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
 \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{1-\alpha/2}(n_1 + n_2 - 2) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (3.29)
 \end{aligned}$$

where

$$S_p = \left\{ \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \right\}^{1/2} \quad (3.30)$$

is the pooled estimator of the common standard deviation $\sigma_1 = \sigma_2 = \sigma$ and where S_1^2 and S_2^2 are given by Equation 3.13. The quantity $t_{1-\alpha/2}(n_1 + n_2 - 2)$ is obtained from Table A5.

If $\sigma_1^2 \neq \sigma_2^2$, an approximate $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\begin{aligned}
 (\bar{X}_1 - \bar{X}_2) - t_{1-\alpha/2}(df) \left[\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^{1/2} \right] \\
 \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{1-\alpha/2}(df) \left[\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^{1/2} \right] \quad (3.31)
 \end{aligned}$$

where df denotes the approximate degrees of freedom given by

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{S_1^4}{n_1^2(n_1 - 1)} + \frac{S_2^4}{n_2^2(n_2 - 1)}} \quad (3.32)$$

This approximation is from Satterthwaite (1946).

A $100(1 - \alpha)\%$ one-sided confidence limit may be obtained from Equation 3.29 or 3.31 by substituting α for $\alpha/2$ and choosing the appropriate side of the two-sided interval.

When σ_1^2 and σ_2^2 are known, Equation 3.31 should be used, with the known values of σ_1^2 and σ_2^2 substituted for S_1^2 and S_2^2 and with $z_{1-\alpha/2}$ substituted for $t_{1-\alpha/2}(df)$.

Example 3.8 Suppose the data in Example 3.5 represent UO_2 containers from two batches produced by different processes. Estimate the difference in true mean UO_2 content of material produced by the two

processes and construct a 95% confidence interval for the difference $\mu_1 - \mu_2$.

Let batch 1 be the X_1 population and batch 2 be the X_2 population and assume that σ_1^2 and σ_2^2 are equal. Equation 3.29 provides the appropriate confidence interval. With $n_1 = 8$, $n_2 = 10$, and $\alpha = 0.05$, the value $t_{0.975}(16) = 2.12$ is obtained from Table A5. Substituting these values with $s_1^2 = (0.016)^2$, $s_2^2 = (0.023)^2$, $\bar{x}_1 = 87.637$, and $\bar{x}_2 = 87.667$ into Equations 3.30 and 3.29 yields

$$s_p = [(8 - 1)(0.016)^2 + (10 - 1)(0.023)^2 / (8 + 10 - 2)]^{1/2} = 0.0202$$

and

$$\begin{aligned} (87.637 - 87.667) - 2.12(0.0202)(1/8 + 1/10)^{1/2} &\leq \mu_1 - \mu_2 \\ &\leq (87.637 - 87.667) + 2.12(0.0202)(1/8 + 1/10)^{1/2} \end{aligned}$$

or

$$-0.050 \leq \mu_1 - \mu_2 \leq -0.010$$

The point estimate of the difference is -0.030% , and the 95% confidence interval is $(-0.050, -0.010)$.

3.5.3 Interval Estimators for Binomial Distribution Parameters

In Section 3.4.2, an unbiased estimator for the binomial parameter p is given by $\hat{p} = Y/n$, where Y is the number of successes observed in a sample of size n . In this section, the sampling distribution of the estimator \hat{p} for large sample sizes is considered, and a confidence interval for p using this sampling distribution is presented. A confidence interval for p for small sample sizes is also considered.

For large sample sizes, the Central Limit Theorem (Section 2.5.2.1) indicates that the quantity

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \quad (3.33)$$

is approximately distributed as $N(0,1)$. That is, for large sample sizes, \hat{p} is approximately distributed as $N[p, p(1-p)/n]$. The discussion in Sec-

tion 2.5.2.2 indicates that this normal approximation to the binomial gives reasonably accurate results for $n > 5/p$ when $p \leq 0.5$ and for $n > 5/(1 - p)$ when $p > 0.5$. Under these conditions on n , the sampling distribution of \hat{p} yields an approximate $100(1 - \alpha)\%$ confidence interval for p , given by

$$\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (3.34)$$

where $\hat{p}(1 - \hat{p})$ is used to estimate $p(1 - p)$.

A one-sided $100(1 - \alpha)\%$ confidence limit is obtained from Equation 3.34 by substituting α for $\alpha/2$ and choosing the appropriate side of the interval.

In practice, the approximate confidence interval computed from Equation 3.34 may yield a lower limit less than zero or an upper limit greater than one. Because p must be between zero and one, the lower or upper confidence limits of the confidence interval should be set to these values if they fall outside them in a particular application.

Example 3.9 For the manual record-keeping system audit problem of Example 3.1, assume that $n = 100$ records are sampled, and $y = 8$ of them are found to contain at least one error. Compute a 95% confidence interval for the population proportion of erroneous records.

The point estimate of p is given by $\hat{p} = y/n = 8/100 = 0.08$. Because $n = 100 > 5/\hat{p} = 62.50$, the normal approximation to the binomial should give satisfactory results. For a 95% confidence interval, $\alpha = 0.05$ and $z_{0.975} = 1.96$. By substituting these quantities into Equation 3.34, the approximate confidence interval is computed as

$$0.08 - 1.96 [0.08 (0.92)/100]^{1/2} \leq p \leq 0.08 + 1.96 [0.08 (0.92)/100]^{1/2}$$

$$0.08 - 0.05 \leq p \leq 0.08 + 0.05$$

$$0.03 \leq p \leq 0.13$$

That is, an approximate 95% confidence interval for p is (0.03, 0.13).

A related confidence interval problem is to determine the sample size required to obtain a $100(1 - \alpha)\%$ confidence interval for p , with half-width w . The sample size is obtained by solving the equation

$$w = z_{1-\alpha/2} [p(1-p)/n]^{1/2}$$

for n , which gives

$$n = z_{1-\alpha/2}^2 \left[\frac{p(1-p)}{w^2} \right] \quad (3.35)$$

Because the sample size is an integer, n is rounded to the nearest integer. When it is desirable to be conservative, n is rounded up.

A maximum value of n is obtained for given w and α when $p = 0.5$. This can be used when no specific prior information is available about the value of p . Although p is unknown, in practice there is usually some knowledge about its possible values. Often this knowledge may be in the form of upper or lower limits on the true unknown probability p . The use of such information with Equation 3.35 is illustrated in the following example.

Example 3.10 A population of 2000 cans of UO_2 powder is to be inspected to verify the gross weights stored in the records system. Each can weighs approximately 10 kg. A can is termed “defective” if its measured weight is not within 20 g of the value in the records system. Calculate the sample size required to obtain a one-sided 90% upper confidence limit with half-width 0.02 for p , the population proportion of defective cans. It is expected that the true value of p is less than 0.05.

Equation 3.35 provides the required approximate sample size where $z_{1-\alpha/2}$ is replaced by $z_{1-\alpha}$ for the one-sided upper confidence limit. Because in Equation 3.35 n is an increasing function of p for fixed α and w when $p \leq 0.5$, a conservative sample size is obtained by substituting the approximate upper limit of 0.05 for p . Then Equation 3.35 yields

$$n = (z_{0.90})^2 \left[\frac{0.05(1-0.05)}{(0.02)^2} \right] = (1.28)^2 (118.75) = 194.56 \cong 195$$

Thus, 195 cans should be sampled to obtain a 90% upper confidence limit with half-width 0.02.

Consider the case of interval estimation of p for smaller sample sizes, where the normal approximation method is not applicable. Recall that the number of successes Y observed in a sample of size n is a binomial ran-

dom variable with mean np and variance $np(1 - p)$. This sampling distribution of Y yields a $100(1 - \alpha)\%$ confidence interval for p of the form

$$p_l \leq p \leq p_u \quad (3.36)$$

where p_u is the value of p such that

$$\sum_{k=0}^y \binom{n}{k} p^k (1 - p)^{n-k} = \frac{\alpha}{2} \quad (3.37)$$

p_l is the value of p such that

$$\sum_{k=y}^n \binom{n}{k} p^k (1 - p)^{n-k} = \frac{\alpha}{2} \quad (3.38)$$

and where y is the number of successes observed in the sample of size n . A one-sided $100(1 - \alpha)\%$ confidence limit is obtained by substituting α for $\alpha/2$ in Equation 3.37 or 3.38. The solution of Equations 3.37 and 3.38 for p_u and p_l is simplified by utilizing the cumulative binomial distribution values from Table A1. The following example illustrates the technique.

Example 3.11 For the manual records-keeping system audit problem of Example 3.9, suppose that $n = 20$ records are sampled and $y = 2$ of them are found to contain at least one error. Develop a 95% confidence interval for the population proportion of erroneous records.

The point estimate of p is given by $\hat{p} = y/n = 2/20 = 0.10$. The sample size of 20 is too small to use the normal approximation method. Hence, the desired confidence interval is obtained from Equations 3.36, 3.37, and 3.38, where $\alpha = 0.05$ and $n = 20$.

Consider first the upper limit p_u , which is the value of p such that

$$\sum_{k=0}^2 \binom{20}{k} p^k (1 - p)^{20-k} = 0.025$$

Entering Table A1 with $n = 20$ and $x = 2$, look for the value 0.025 in the body of the table; the corresponding value of p from the upper margin is p_u . Usually the specified value of $\alpha/2$ (or α for a one-sided interval) does not appear in Table A1. The bounding values and their associated p values are used with linear interpolation to approximate the value of p associated with the chosen value of $\alpha/2$ (or α). From Table A1 with

$n = 20$ and $y = 2$, note that $0.30 < p_u < 0.35$. Linear interpolation yields

$$p_u = 0.30 + \frac{0.0355 - 0.025}{0.0355 - 0.0121} (0.35 - 0.30) = 0.322$$

Now consider the lower limit p_l , which is the value of p such that

$$\sum_{k=2}^{20} \binom{20}{k} p^k (1-p)^{20-k} = 0.025$$

This may be solved using the cumulative binomial probabilities in Table A1 by noting that

$$P(Y \geq y) = 1 - P(Y \leq y - 1)$$

To make the cumulative binomial Table A1 directly applicable, Equation 3.38 can be rewritten as

$$\sum_{k=0}^{y-1} \binom{n}{k} p^k (1-p)^{n-k} \cong 1 - \frac{\alpha}{2}$$

Then, entering Table A1 with $n = 20$ and $x = 2 - 1 = 1$, note that $0 < p_l < 0.05$. Using linear interpolation gives

$$p_l = 0 + \frac{1.0 - 0.975}{1.0 - 0.7348} (0.05 - 0.0) = 0.005$$

Thus, a 95% confidence interval for the population proportion of erroneous records is (0.005, 0.322).

Other methods for constructing confidence intervals for the binomial parameter p are found in Nelson (1982); however, they are not discussed or illustrated here.

Consider two binomial random variables X_1 and X_2 for two populations with parameters p_1 and p_2 . Suppose that independent random samples of size n_1 and n_2 are available. If n_1 and n_2 both satisfy the "large sample" conditions for applying the normal approximation to the binomial, an approximate $100(1 - \alpha)$ confidence interval for the difference $p_1 - p_2$ is given by

$$\begin{aligned}
 (\hat{p}_1 - \hat{p}_2) - z_{1-\alpha/2} \sqrt{\text{Vâr}(\hat{p}_1 - \hat{p}_2)} &\leq p_1 - p_2 \\
 &\leq (\hat{p}_1 - \hat{p}_2) + z_{1-\alpha/2} \sqrt{\text{Vâr}(\hat{p}_1 - \hat{p}_2)} \quad (3.39)
 \end{aligned}$$

where

$$\text{Vâr}(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2} \quad (3.40)$$

An approximate $100(1 - \alpha)\%$ one-sided confidence limit is obtained from Equation 3.39 by substituting α for $\alpha/2$ and choosing the appropriate side of the two-sided interval.

3.5.4 Interval Estimators for the Hypergeometric Distribution Parameter D

For finite populations where sampling is performed without replacement, the hypergeometric distribution (Section 2.4.1) is applicable. The maximum likelihood estimator of D, the number of defectives in the population, is given in Table 3.1 of Section 3.4.3. A confidence interval for D is now considered.

Assume that a random sample of n items is selected from a population of N items which contains D defectives. The number of defectives Y observed in a sample of size n has a hypergeometric distribution. This sampling distribution of Y yields a $100(1 - \alpha)\%$ confidence interval for D of the form

$$D_l \leq D \leq D_u \quad (3.41)$$

where D_u is the smallest integer value of D such that

$$\sum_{k=0}^y \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}} \leq \frac{\alpha}{2} \quad (3.42)$$

and D_l is the largest integer value of D such that

$$\sum_{k=y}^n \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}} \leq \frac{\alpha}{2} \quad (3.43)$$

and where y is the number of defectives observed in the sample of size n . A one-sided confidence limit for D is obtained by choosing the appropriate side of Equation 3.41 and then substituting α for $\alpha/2$ in Equation 3.42 or 3.43.

In most practical applications, the number of observed defectives y is small, and Equations 3.42 and 3.43 can be solved by trial and error using a hand calculator. A programmable calculator or computer is recommended, however, for larger values of y . Values of D_u for one-sided upper confidence limits have been tabulated by Stewart (1973). Note that it is usually not possible to find values of D_u and D_l such that Equations 3.42 and 3.43 are exactly equal to $\alpha/2$. Thus, the actual confidence level is $(1 - \alpha^*)$, where α^* is the sum of the actual values computed from Equations 3.42 and 3.43 when the chosen values of D_u and D_l are substituted for D .

Example 3.12 This example is adapted from Jaech (1973). A population of $N = 300$ fuel rods is to be sampled, and the sampled rods will be scanned to verify the ^{235}U enrichment. A sample of size $n = 50$ rods is scanned, and all are verified as having the enrichment stated on the inventory listing. Find the upper 95% confidence limit for the number of defectives D in the population of 300 rods.

Equation 3.42, with $\alpha/2$ replaced by $\alpha = 0.05$, provides the desired limit. Here $N = 300$, $n = 50$, the number of observed defectives in the sample is $y = 0$, and it is necessary to find D_u , the smallest value of D such that

$$F_Y(0) = \frac{\binom{D}{0} \binom{300-D}{50-0}}{\binom{300}{50}} \leq 0.05$$

For illustration, $F_Y(0)$ is evaluated for D ranging from 12 to 20 with the following results

D	$F_Y(0)$	D	$F_Y(0)$
20	0.0228	15	0.0604
19	0.0278	14	0.0731
18	0.0338	13	0.0886
17	0.0410	12	0.1072
16	0.0498		

$D_u = 16$ is the smallest value of D such that $F_Y(0) \leq 0.05$. Therefore, the upper 95% confidence limit for D is $D_u = 16$.

3.6 ESTIMATORS FOR FUNCTIONS OF RANDOM VARIABLES

Population means and variances of functions of random variables are considered in Section 2.7. In both Sections 2.7.1 and 2.7.2, the expressions for the mean and variance of a function involve the means, variances, and covariances of the individual random variables in the function. In practice, however, the true values of these parameters are often unknown, and they must be estimated from sample data. The sample estimates are then substituted for the unknown parameters to compute point estimates of the mean and variance of the function.

In this section, point estimators are presented for the mean and variance of a function of random variables (Section 3.6.1) and interval estimators for the mean of a function of random variables (Section 3.6.2).

3.6.1 Point Estimators

In Section 2.7.1, Equations 2.122 and 2.123 are the population mean and variance of a linear combination of random variables of the form

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k \quad (3.44)$$

Point estimators for the mean and variance of the random variable Y are obtained by replacing the unknown population means, variances, and covariances of the X_i in Equations 2.122 and 2.123 with their point estimators, which gives

$$\hat{\mu}_Y = a_0 + a_1 \hat{\mu}_1 + a_2 \hat{\mu}_2 + \dots + a_k \hat{\mu}_k \quad (3.45)$$

and

$$\hat{\sigma}_Y^2 = \sum_{i=1}^k a_i^2 \hat{\sigma}_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j>i} a_i a_j \hat{\sigma}_{ij} \quad (3.46)$$

If X_1, X_2, \dots, X_k are mutually independent, all of the $\hat{\sigma}_{ij}$ are set equal to zero, and Equation 3.46 reduces to

$$\hat{\sigma}_Y^2 = \sum_{i=1}^k a_i^2 \hat{\sigma}_i^2 \quad (3.47)$$

In Equations 3.45, 3.46, and 3.47, the estimators $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ for the mean and variance of X_i are often the sample mean \bar{X}_i and the sample variances S_i^2 given by Equations 3.11 and 3.13, respectively. The estimator $\hat{\sigma}_{ij}$ is often the sample covariance of X_i and X_j . That is, when n paired observations (Section 4.4) are taken on any two random variables X_i and X_j , the sample covariance is given by

$$S_{ij} = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{n - 1}$$

$$= \frac{\sum_{k=1}^n (X_{ik} X_{jk}) - \left(\sum_{k=1}^n X_{ik} \right) \left(\sum_{k=1}^n X_{jk} \right) / n}{n - 1} \quad (3.48)$$

In many specialized applications, however, such as regression (presented in Chapter 6), the means, variances, and covariances of the random variables in a function are estimated by other methods. Thus, the notation $\hat{\mu}_i$, $\hat{\sigma}_i^2$, and $\hat{\sigma}_{ij}$ indicates only that estimators of the mean, variance, and covariance are to be used, but no particular estimators are implied.

Example 3.13 Suppose that in Example 2.10, the same scale is used to weigh both the plugs and springs but that another scale is used to weigh both the cladding and the fuel. This implies that a covariance exists between the observed plug and spring weights and also between the observed cladding and fuel weights.

Assume that the means, variances, and covariances of the four random variables are unknown but that point estimates have been calculated from sample data. For illustration, assume that the point estimates of the means and variances have the same values as the parameters given in Example 2.10. These are repeated below in the point estimate notation.

Component	Sample mean weight, in grams	Sample variance, in grams ²
Cladding: X_1	$\bar{x}_1 = 554.89$	$s_1^2 = 16.8100$
Plug: X_2	$\bar{x}_2 = 10.64$	$s_2^2 = 0.0016$
Spring: X_3	$\bar{x}_3 = 16.07$	$s_3^2 = 0.0036$
Fuel: X_4	$\bar{x}_4 = 1814.85$	$s_4^2 = 6.0025$

Assume also that the covariance estimates, $\hat{\sigma}_{14} = 0.9041$ and $\hat{\sigma}_{23} = 0.0004$, are computed from sample data. Compute the point estimates of the mean, variance, and standard deviation of Y , the total weight of a fuel rod.

Applying Equations 3.45 and 3.46, the estimates of μ_Y and σ_Y^2 are

$$\hat{\mu}_Y = 554.89 + 10.64 + 16.07 + 1814.85 = 2396.45 \text{ g}$$

and

$$\begin{aligned} \hat{\sigma}_Y^2 &= 16.8100 + 0.0016 + 0.0036 + 6.0025 \\ &\quad + 2[(0.9041) + (0.0004)] \\ &= 24.6267 \text{ g}^2 \end{aligned}$$

The estimate of σ_Y is then

$$\sigma_Y = \sqrt{24.6267} = 4.96 \text{ g}$$

Equations 2.126 and 2.127 of Section 2.7 are approximate expressions for the population mean and variance of an arbitrary function

$$Y = h(X_1, X_2, \dots, X_k) \quad (3.49)$$

of random variables X_1, X_2, \dots, X_k . Point estimators for the approximate mean and variance of Y are obtained by replacing the unknown population means, variances, and covariances of the X_i in Equations 2.126 and 2.127 with their point estimators. This gives

$$\hat{\mu}_Y = h(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k) \quad (3.50)$$

and

$$\hat{\sigma}_Y^2 = \sum_{i=1}^k \left(\frac{\partial h}{\partial X_i} \right)^2 \hat{\sigma}_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j>i} \left(\frac{\partial h}{\partial X_i} \right) \left(\frac{\partial h}{\partial X_j} \right) \hat{\sigma}_{ij} \quad (3.51)$$

where the partial derivatives $(\partial h / \partial X_i)$ are evaluated at $X_i = \hat{\mu}_i$. If the random variables X_1, X_2, \dots, X_k are mutually independent, Equation 3.51 reduces to

$$\hat{\sigma}_Y^2 = \sum_{i=1}^k \left(\frac{\partial h}{\partial X_i} \right)^2 \hat{\sigma}_i^2 \quad (3.52)$$

As in the discussion of Equations 3.45, 3.46, and 3.47, the notation $\hat{\mu}_i$, $\hat{\sigma}_i^2$, and $\hat{\sigma}_{ij}$ in Equations 3.50, 3.51, and 3.52 indicates that point estimators are to be used, but no particular estimators are implied.

When Equations 3.50 and 3.51 are applied to obtain estimates of μ_Y and σ_Y^2 , the validity of the resulting estimates depends upon the effects of two primary factors, as discussed in Section 2.7.2. First, the first-order Taylor series approximation may be a poor representation of the function $h(X_1, X_2, \dots, X_k)$. Second, estimating the partial derivatives, variances, and covariances in Equation 2.127 by substituting sample estimates for μ_i , σ_i^2 , and σ_{ij} may result in estimates with poor statistical properties. The combined effect of these two factors may result in poor estimates of inadequate approximations.

Jaech (1973) points out that in most nuclear material control applications the errors introduced by substituting sample estimates for the unknown parameters overshadow the errors that might be introduced by an inadequate approximation of the function $h(X_1, X_2, \dots, X_k)$. Thus, there is usually little need to be concerned about the adequacy of the approximation. Generally, if the approximation is deemed reasonable over the relevant range of the X_i variables, and if each $\hat{\sigma}_i$ is small relative to $\hat{\mu}_i$, the estimates obtained by applying Equations 3.50 and 3.51 are acceptable for practical use.

3.6.2 Interval Estimators for Means

Often, in nuclear material accounting applications, an interval estimate for the mean μ_Y of a random variable Y is desired, where the random variable is some function of other random variables (which are often measured quantities). If Y is a linear combination, as given by Equation 3.44, where the X_i are independently distributed as $N(\mu_i, \sigma_i^2)$, then from Section 2.7.1, $Y \sim N(a_0 + \sum a_i \mu_i, \sum a_i^2 \sigma_i^2)$. When $\hat{\mu}_i = \bar{X}_i$ is used, then $\hat{\mu}_Y$ given by Equation 3.45 is distributed as $N(a_0 + \sum a_i \mu_i, \sum a_i^2 \sigma_i^2/n_i)$.

Even when the X_i are not normally distributed, the sample means \bar{X}_i are closer to being normally distributed than the X_i , and a linear combination of the sample means can often be assumed to be approximately normally distributed. Suppose that random samples of size n_i are taken from each X_i population and yield point estimates \bar{x}_i and s_i^2 for the mean and variance of each X_i . Based upon these point estimates and the sampling distribution of $\hat{\mu}_Y$, a $100(1 - \alpha)\%$ confidence interval for μ_Y is given by

$$\hat{\mu}_Y - t_{1-\alpha/2}(\nu) \left[\sum_{i=1}^k (a_i^2 s_i^2/n_i) \right]^{1/2} \\ \leq \mu_Y \leq \hat{\mu}_Y + t_{1-\alpha/2}(\nu) \left[\sum_{i=1}^k (a_i^2 s_i^2/n_i) \right]^{1/2} \quad (3.53)$$

where $\hat{\mu}_Y$ is given by Equation 3.45, $t_{1-\alpha/2}(v)$ is obtained from Table A5, and v is the approximate degrees of freedom obtained from Satterthwaite's (1946) formula

$$v = \frac{\left(\sum_{i=1}^k a_i^2 s_i^2 / n_i \right)^2}{\sum_{i=1}^k \left(a_i^2 s_i^2 / n_i \right)^2 / (n_i - 1)} \quad (3.54)$$

If $v \geq 30$, then $z_{1-\alpha/2}$ from Table A3 may be used instead of $t_{1-\alpha/2}(v)$ in Equation 3.53 with satisfactory results.

Example 3.14 In Example 2.10, assume that the parameter values are actually point estimates of the means and variances of the rod components obtained from samples of size $n_1 = 10$, $n_2 = 20$, $n_3 = 15$, and $n_4 = 8$. Further, assume that the four components are independently normally distributed. Compute a 95% confidence interval for the mean total weight of a fuel rod.

Noting that $a_0 = 0$ and all $a_i = 1$, the sample estimates are substituted into Equation 3.45 to compute

$$\hat{\mu}_Y = 554.89 + 10.64 + 16.07 + 1814.85 = 2396.45 \text{ g}$$

The variance of $\hat{\mu}_Y$ is estimated by the quantity

$$\begin{aligned} \text{Var}(\hat{\mu}_Y) &= \sum_{i=1}^4 \frac{a_i^2 s_i^2}{n_i} = \frac{16.8100}{10} + \frac{0.0016}{20} + \frac{0.0036}{15} + \frac{6.0025}{8} \\ &= 2.4316 \end{aligned}$$

Substituting the variance estimates and sample sizes into Equation 3.54 gives the approximate degrees of freedom

$$\begin{aligned} v &= \frac{[(16.81/10) + (0.0016/20) + (0.0036/15) + (6.0025/8)]^2}{\frac{(16.81/10)^2}{9} + \frac{(0.0016/20)^2}{19} + \frac{(0.0036/15)^2}{14} + \frac{(6.0025/8)^2}{7}} \\ &= \frac{5.9128}{0.3944} = 14.99 \cong 15 \end{aligned}$$

Then, entering Table A5 with 15 df, the value $t_{0.975}(15) = 2.13$ is obtained. A 95% confidence interval for μ_Y is calculated from Equation 3.53 as

$$2396.45 - 2.13\sqrt{2.4316} \leq \mu_Y \leq 2396.45 + 2.13\sqrt{2.4316}$$

$$2396.45 - 3.32 \leq \mu_Y \leq 2396.45 + 3.32$$

$$2393.13 \leq \mu_Y \leq 2399.77$$

That is, a 95% confidence interval for the mean total fuel rod weight is (2393.13 g, 2399.77 g).

3.7 TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

This section presents techniques for constructing **one-sided** and **two-sided statistical tolerance** limits for a normal distribution with unknown mean and variance. The techniques are based upon the assumption that a random sample of size $n > 1$ has been taken from a population with a pdf that is either normal or very nearly so. For cases where the normality assumption is in doubt, methods for constructing distribution-free tolerance limits such as those presented by Natrella (1966) and Conover (1980) should be considered.

3.7.1 Two-Sided Tolerance Limits

The confidence interval given by Equation 3.26 provides an interval estimate for the unknown mean. Tolerance limits differ from confidence intervals in that tolerance limits provide an interval within which at least a specified proportion P of the population lies, with specified probability $1 - \alpha$. To introduce the notion of tolerance limits, consider first a random variable $X \sim N(\mu, \sigma^2)$, where the values of μ and σ^2 are known. Suppose that it is necessary to find limits, L and U , such that the interval (L, U) includes $100P\%$ of the possible values of X . This problem can be solved by the methods presented in Section 2.5.2. For example, if $P = 0.95$, then the limits are $L = \mu - z_{0.975} \sigma$ and $U = \mu + z_{0.975} \sigma$; that is, 95% of the possible values of X are in the interval given by $\mu \pm 1.96 \sigma$. In practice, however, μ and σ are usually unknown and must be estimated. Because the estimators \bar{X} and S^2 are random variables, substituting the point estimates \bar{x} and s for μ and σ in the above formulas for L and U ignores the sampling distributions of \bar{X} and S^2 and can yield misleading results.

Statistical tolerance limits are computed from a procedure that takes into account the sampling distributions of \bar{X} and S^2 . Consider a random variable $X \sim N(\mu, \sigma^2)$, where the values of μ and σ^2 are unknown. Suppose that an interval (X_L, X_U) must be constructed in such a way that we are $100(1-\alpha)\%$ confident that at least $100P\%$ of the possible values of X are inside of the interval. Assume that a random sample of size n is taken and that point estimates of μ and σ^2 are computed from Equations 3.11 and 3.13. Tolerance limits with specified values of $1-\alpha$ and P are given by

$$X_L = \bar{X} - K(n, 1-\alpha, P) S \quad (3.55)$$

and

$$X_U = \bar{X} + K(n, 1-\alpha, P) S \quad (3.56)$$

where

$$K(n, 1-\alpha, P) = z_{1-(1-P)/2} \left[\frac{(n-1)}{\chi_{\alpha}^2(n-1)} \right]^{1/2} \left(1 + \frac{1}{2n} \right) \quad (3.57)$$

and where $z_{1-(1-P)/2}$ is obtained from Table A3 and $\chi_{\alpha}^2(n-1)$ is obtained from Table A4.

The derivation of Equation 3.57 is presented by Hald (1952), who also provides a table of values of $K(n, 1-\alpha, P)$ for selected values of $1-\alpha$ and P with n ranging from 4 to ∞ in selected increments. More extensive tables of $K(n, 1-\alpha, P)$ are provided by Natrella (1966).

Example 3.15 Using the data from Example 3.4, construct tolerance limits such that we can be 90% confident that at least 95% of the fuel rod weights in the population are within the tolerance limits.

The point estimates of μ and σ are $\bar{x} = 2394.43$ g and $s = 4.44$ g. With $P = 0.95$, the value $z_{0.975} = 1.96$ is obtained from Table A3. With $1-\alpha = 0.90$ and $n-1 = 11$, the value $\chi_{0.10}^2(11) = 5.58$ is obtained from Table A4. Substituting these values into Equation 3.57 yields

$$K(12, 0.90, 0.95) = 1.96 \sqrt{\frac{11}{5.58}} \left(1 + \frac{1}{24} \right) = 2.87$$

Application of Equations 3.55 and 3.56 yields the tolerance limits

$$X_L = 2394.43 - 2.87(4.44) = 2381.69$$

and

$$X_U = 2394.43 + 2.87(4.44) = 2407.17$$

Thus, we can be 90% confident that at least 95% of the fuel rod weights in this facility are between 2381.69 g and 2407.17 g.

3.7.2 One-Sided Tolerance Limits

For some applications, interest is focused on estimating a single limit above (or below) which at least a specified proportion P of the population lies with specified probability $1 - \alpha$. In such cases, either an upper or lower one-sided statistical tolerance limit is appropriate. Assume that a random sample of size n has been taken from a normal distribution with unknown mean and variance and that point estimates of μ and σ^2 are computed from Equations 3.11 and 3.13. For specified values of $1 - \alpha$ and P , a lower tolerance limit is given by

$$X_L^* = \bar{X} - K^*(n, 1 - \alpha, P)S \quad (3.58)$$

and an upper tolerance limit is given by

$$X_U^* = \bar{X} + K^*(n, 1 - \alpha, P)S \quad (3.59)$$

The asterisk notation is used to emphasize that $K^*(n, 1 - \alpha, P)$ is *not* the same as $K(n, 1 - \alpha, P)$ given by Equation 3.57. $K^*(n, 1 - \alpha, P)$ is given by

$$K^*(n, 1 - \alpha, P) = \frac{z_P + \sqrt{z_P^2 - ab}}{a} \quad (3.60)$$

where

$$a = 1 - \frac{z_{1-\alpha}^2}{2(n-1)} \quad (3.61)$$

and

$$b = z_P^2 - \frac{z_{1-\alpha}^2}{n} \quad (3.62)$$

The values of $z_{1-\alpha}$ and z_p are obtained from Table A3. Equations 3.58 through 3.62 are from Natrella (1966), where extensive tables of $K^*(n, 1 - \alpha, P)$ are also provided.

Example 3.16 Using the data from Example 3.4, construct an upper tolerance limit X_U^* such that we can be 90% confident that at least 95% of the fuel rod weights in the population are below X_U^* .

The point estimates of μ and σ are $\bar{x} = 2394.43$ g and $s = 4.44$ g. With $1 - \alpha = 0.90$ and $P = 0.95$, the values $z_{0.90} = 1.29$ and $z_{0.95} = 1.65$ are obtained from Table A3. Application of Equations 3.61 and 3.62 yields

$$a = 1 - \frac{(1.29)^2}{2(11)} = 0.924$$

and

$$b = (1.65)^2 - \frac{(1.29)^2}{12} = 2.584$$

and substituting these values into Equation 3.60 yields

$$K^*(12, 0.90, 0.95) = \frac{1.65 + \sqrt{(1.65)^2 - (0.924)(2.584)}}{0.924} = 2.41$$

Finally, application of Equation 3.59 yields

$$X_U^* = 2394.43 + 2.41(4.44) = 2405.13 \text{ g}$$

Thus, we can be 90% confident that at least 95% of the fuel rod weights in the population are less than 2405.13 g.

REFERENCES

- Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, 1972, *Robust Estimates of Location*, Princeton University Press, Princeton, New Jersey.
- Conover, W. J., 1980, *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York.
- DeGroot, M. H., 1970, *Optimal Statistical Decisions*, McGraw-Hill, New York.
- Graybill, F. A., 1976, *Theory and Application of the Linear Model*, Duxbury Press, Boston.

- Hahn, G. J., and S. S. Shapiro, 1967, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York.
- Hald, A., 1952, *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York.
- Hoel, P. G., S. C. Port, and C. J. Stone, 1970, *Introduction to Statistical Theory*, Houghton Mifflin Co., Boston.
- Hogg, R. V., and A. T. Craig, 1970, *Introduction to Mathematical Statistics*, 3rd Ed., Macmillan Publishing Co., Inc., New York.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Johnson, N. L., and S. Kotz, 1969, *Discrete Distributions*, John Wiley & Sons, Inc., New York.
- , and S. Kotz, 1970, *Continuous Univariate Distributions*, Vols. 1 and 2, John Wiley & Sons, Inc., New York.
- Kendall, M. G., and A. Stuart, 1973, *The Advanced Theory of Statistics*, Vol. 2, Hafner Publishing Co., New York.
- Larson, H. J., 1969, *Introduction to Probability Theory and Statistical Inference*, John Wiley & Sons, Inc., New York.
- Lindgren, B. W., 1976, *Statistical Theory*, 3rd Ed., Macmillan Publishing Co., Inc., New York.
- Martz, H. F., and R. A. Waller, 1982, *Bayesian Reliability Analysis*, John Wiley & Sons, Inc., New York.
- Natrella, M. G., 1966, *Experimental Statistics*, National Bureau of Standards Handbook 91, Second Printing, U. S. Government Printing Office, Washington, D. C.
- Nelson, W., 1982, *Applied Life Data Analysis*, John Wiley & Sons, Inc., New York.
- Rohatgi, V. K., 1984, *Statistical Inference*, John Wiley & Sons, Inc., New York.
- Satterthwaite, F. E., 1946, An Approximate Distribution of Estimates of Variance Components, *Bio. Bull.*, 2: 110-114.
- Stewart, K. B., 1973, *Upper Confidence Limit Tables for the Hypergeometric Distribution*, BNWL-1681, Battelle Pacific Northwest Laboratories, Richland, Washington.

CHAPTER 4

Hypothesis Testing

4.0 INTRODUCTION

In nuclear material control, situations frequently arise that require deciding whether or not it is reasonable to continue operating under a specified assumption about the value of one or more population or process parameters. A formal statement that specifies the assumed parameter values is called a **hypothesis**, and the procedure for deciding whether or not it is reasonable to continue operating under this hypothesis is called **hypothesis testing**. This is an extremely useful aspect of statistical inference because many types of decision problems can be formulated as hypotheses about population or process parameters. This chapter presents fundamental concepts and procedures for some important applications of hypothesis testing.

4.1 BASIC CONCEPTS

This section gives a brief overview of hypothesis testing. Some basic concepts and terms are introduced to facilitate subsequent discussions.

4.1.1 Statistical Hypotheses

A statistical hypothesis is a formal statement about the probability density function of one or more random variables associated with a population or process. This statement is usually expressed in terms of the parameters of the pdf; many of the statistical hypotheses encountered in material control applications specify values for population or process parameters, such as means, differences between means, variances, proportions, and so on. For example, suppose that it is of interest to estimate the uranium content of a particular batch of UO_2 powder. The normal practice is to multiply the net weight of UO_2 powder by a standard factor of 87.6%. Suppose that in this case, however, there is some reason to suspect the applicability of the standard factor and it is of interest to decide whether or not the true mean percent uranium (say μ) for this batch is 87.6%. This can be expressed formally as the set of hypotheses

$$H_0: \mu = 87.6\%$$

vs.

(4.1)

$$H_A: \mu \neq 87.6\%$$

The statement $H_0: \mu = 87.6\%$ is called the **null hypothesis**, and the statement $H_A: \mu \neq 87.6\%$ is called the **alternative hypothesis**. In this example, the alternative hypothesis includes values of μ that are either greater than or less than 87.6%.

We are interested in deciding whether or not it is reasonable to continue operating under the specified null hypothesis. Hypothesis testing leads to such a decision based upon an evaluation of the information in a random sample from the population or process of interest. If the sample information indicates that the random sample could reasonably have come from a distribution with the parameter value(s) specified in the null hypothesis, then a decision is typically made to continue operating as though the null hypothesis were true. If, however, the indication is that the observed sample results would have a small probability of occurring under the null hypothesis, then we conclude that the null hypothesis is probably false, which implies that the alternative hypothesis is probably true. This could indicate a need for investigative studies or process modifications before operations continue.

It is important to remember that hypotheses are *always* statements about the population under study and not statements about the sample data. On this subject, Jaech (1973) has the following comment:

A misapplication of hypothesis testing occurs when the hypotheses are formulated after the data are collected. This practice of using data to suggest hypotheses is at variance with the whole idea of hypothesis testing. Once the data have been collected, anyone with an ounce of brains and a lesser amount of integrity, or more charitably stated, "any misguided individual," can formulate hypotheses that will be rejected when tested against the data. The statement is then made that "the data prove that such and such. . . ." It is an acceptable practice, of course, to formulate hypotheses based on given data sets, but it is simply not acceptable to test their validity with the same set of data that provided the basis for formulating the hypotheses in the first place.

It should be noted that, although Jaech's comment is definitely relevant to the topics covered in this chapter, there are special cases (primarily associated with comparative studies) in which it is legitimate to test hypotheses that have been suggested by the data. Specialized techniques, however, such as the one presented in Section 5.1.4.4., must be used to control the risk of reaching erroneous conclusions.

The value of a population parameter specified in the null hypothesis ($\mu = 87.6\%$ in the above example) is usually determined in one of three ways. First, it may result from experience or knowledge about the process under study or from prior experimentation. In this case, the objective of hypothesis testing is to determine whether the process or experimental con-

ditions have changed. Second, a parameter value may be determined from a scientific theory or model pertaining to the process under study. Here the objective of hypothesis testing is to verify the theory or model. Third, a parameter value may be determined by external considerations, such as design or engineering specifications, contractual obligations, or imposed regulations. In this case, the objective of hypothesis testing is to assess compliance.

The hypothesis testing procedures covered in this chapter require that a **test statistic** be computed from the sample data. This test statistic is then compared with a prespecified table value to make a decision about the null hypothesis. For example, consider testing the null hypothesis in Equation 4.1 about the true mean percent uranium in a batch of UO_2 powder. Suppose that five powder samples are randomly selected from the batch of interest. Each powder sample is analyzed to determine its percent uranium. The five percent uranium values are then averaged, giving a sample mean \bar{X} , which might be used as a test statistic as follows: If $\bar{X} \leq 87.57$ or if $\bar{X} \geq 87.63$, we conclude that the true percent uranium for this particular batch of powder is probably not 87.6. That is, we **reject** the null hypothesis $H_0: \mu = 87.6\%$. Rejecting H_0 implies that the alternative hypothesis H_A is probably true. Thus, the set of possible values of \bar{X} that are less than 87.57 or greater than 87.63 is called the **rejection region** or **critical region** for the test. On the other hand, if $87.57 < \bar{X} < 87.63$, then we *fail to reject* the null hypothesis. Note that the values 87.57% and 87.63% were arbitrarily chosen for this example. The details of constructing an appropriate test statistic and determining the rejection region are presented in subsequent sections for specific situations.

4.1.2 Type I and Type II Errors

The decision of whether or not to reject a null hypothesis is based upon information from a random sample and thus is subject to error. There are two possible types of decision errors that can be made when testing hypotheses. If a null hypothesis is rejected when in fact it is true, a **Type I error** is made. If a null hypothesis is not rejected when in fact it is false, a **Type II error** is made. The possible outcomes of a hypothesis test are summarized in Table 4.1.

TABLE 4.1

Possible Outcomes of Hypothesis Testing

Decision	True situation	
	H_0 is true	H_0 is false
Do not reject H_0	No error	Type II error
Reject H_0	Type I error	No error

It is common practice in statistics to let α designate the probability of making a Type I error and to let β designate the probability of making a Type II error. These conditional probabilities are expressed formally as

$$\alpha = \Pr(\text{Type I error}) = \Pr(\text{reject } H_0 \mid H_0 \text{ is true}) \quad (4.2)$$

and

$$\beta = \Pr(\text{Type II error}) = \Pr(\text{do not reject } H_0 \mid H_0 \text{ is false}) \quad (4.3)$$

In practice, α is commonly called the **significance level** of the test. The probability of correctly rejecting a false null hypothesis is called the **power** of the test and is formally given by

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \mid H_0 \text{ is false}) \quad (4.4)$$

It is important to note that because a hypothesis test is based upon sample data which generally do not conclusively rule out either the null or the alternative hypothesis, α and β are greater than zero in most applications. Thus, failing to reject a null hypothesis *does not prove* it is true, and rejecting a null hypothesis *does not prove* it is false. It may be possible, however, to design an experiment or sampling plan and hypothesis testing procedure that fixes α and β at suitably small values.

The definition of “suitably small” values of α and β depends upon the consequences of making a Type I or a Type II error. For example, in a criminal trial, the null hypothesis is that the accused person is innocent. A Type I error results in punishing an innocent person, whereas a Type II error results in setting a criminal free. In this situation, a Type I error is usually considered by society to be more serious. A parachute manufacturer, however, may want to test the null hypothesis that each of the parachutes he sells will open when needed. In this case, a Type II error results in selling a defective parachute, which could have far more serious consequences than a Type I error of choosing not to sell a good parachute.

One nuclear material accounting application involves testing the null hypothesis that no nuclear material was lost or diverted from a given plant or area of a plant during a specified period. In this application, a Type I error is commonly called a **false alarm**, and α is referred to as the **false alarm rate**. This is because a Type I error is made by concluding that there was a loss or diversion when in fact there was not. Such a false alarm could result in an unnecessary investigation and inspection of the facility. On the other hand, a Type II error is made by concluding that there was no loss or diversion when in fact there was. This is almost

always the more serious error. It is clear, however, that small values for both α and β are desirable in such situations. Considering the trade-offs between the Type I and Type II errors from a regulatory point of view, however, a small value of β is more desirable than a small value of α when both α and β cannot simultaneously be made small.

4.2 SINGLE SAMPLE FROM A NORMAL DISTRIBUTION

In this section, methods are presented for testing hypotheses about the parameters of a normal distribution. The first case involves testing hypotheses about the mean μ when the variance σ^2 is known. The procedure is then generalized to the case of testing hypotheses about the mean μ when σ^2 is unknown and is estimated from the data. Finally, a method for testing hypotheses about σ^2 is presented.

4.2.1 Tests About the Mean, Variance Known

This section presents methods for testing two-sided and one-sided hypotheses about the unknown mean of a normal distribution when the variance is known. These methods are based upon the sample mean \bar{X} , where the random sample is assumed to be from a $N(\mu, \sigma^2)$ distribution. Thus, $\bar{X} \sim N(\mu, \sigma^2/n)$, where n is the sample size. Recall from the Central Limit Theorem (Section 2.5.2.1) that for distributions with mean μ and finite variance σ^2 , \bar{X} has approximately a $N(\mu, \sigma^2/n)$ distribution, provided the sample size n is sufficiently large. Thus, the methods of this section are applicable for making approximate hypothesis tests about means from many distributions. However, for a given distribution, the validity of the approximate test results depends upon how closely $N(\mu, \sigma^2/n)$ approximates the true distribution of \bar{X} (see the rules of thumb for applying the Central Limit Theorem in Section 2.5.2.1).

4.2.1.1 Two-Sided Tests

Consider a random variable X which is a measure of a characteristic of some process or population of interest. Assume that X has a $N(\mu, \sigma^2)$ distribution with μ unknown but σ^2 known. The hypothesis to be tested is

$$H_0: \mu = \mu_0$$

against the alternative

(4.5)

$$H_A: \mu = \mu_A \neq \mu_0$$

where μ_0 is a specified value, and μ_A is an alternative value which is unspecified. This form of H_A is called a **two-sided alternative hypothesis**

because it includes values of μ which are either greater than or less than μ_0 .

Assume that a random sample of size n is drawn from the population of interest, and the sample mean \bar{X} is computed. If the null hypothesis is true, $\bar{X} \sim N(\mu_0, \sigma^2/n)$. Thus, if a value of \bar{X} is observed which is quite different from μ_0 , this indicates that H_0 may not be true. The size of the difference between \bar{X} and μ_0 is usually expressed in units of σ/\sqrt{n} , the standard deviation of \bar{X} . This quantity is called the **test statistic** and is given by

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad (4.6)$$

Referring to Section 3.5.2, and in particular to Equation 3.23, Z has the $N(0, 1)$ distribution when H_0 is true.

Consequently, if H_0 is true, the probability is $1 - \alpha$ that an observed value of the test statistic Z will fall between $-z_{1-\alpha/2}$ and $z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the value from the $N(0, 1)$ distribution such that $\Pr(Z \leq z_{1-\alpha/2}) = 1 - \alpha/2$. When H_0 is true, the probability is α that a value of Z will fall in the region $Z \leq -z_{1-\alpha/2}$ or $Z \geq z_{1-\alpha/2}$. This is illustrated in Fig. 4.1.

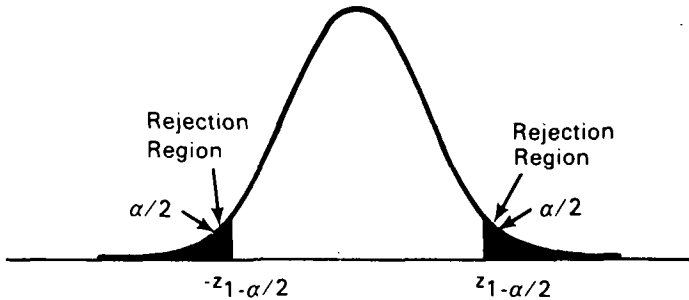


Figure 4.1 Distribution of Z when $H_0: \mu = \mu_0$ is true.

The value of α is specified in advance of the test and is the probability of making a Type I error. The value of $z_{1-\alpha/2}$ is obtained from Table A3 by entering the body of the table, locating $1 - \alpha/2$, and then reading $z_{1-\alpha/2}$ from the margins of the table. For example, if $\alpha = 0.05$, then $1 - \alpha/2 = 0.975$. Entering Table A3, the value of Z for which $\Phi(Z) = 0.975$ is $z_{0.975} = 1.96$.

The hypothesis test proceeds by comparing Z with $\pm z_{1-\alpha/2}$. H_0 is rejected if

$$Z \leq -z_{1-\alpha/2}$$

or (4.7)

$$Z \geq z_{1-\alpha/2}$$

H_0 is not rejected if

$$-z_{1-\alpha/2} < Z < z_{1-\alpha/2} \quad (4.8)$$

The region defined by Equation 4.7 is called the **rejection region** for the test, and the quantities $-z_{1-\alpha/2}$ and $z_{1-\alpha/2}$ are called **critical values** or **rejection limits**. If H_0 is rejected, \bar{X} is said to be significantly different than μ_0 at the α level of significance. This is sufficient evidence to conclude that H_0 is probably false, or equivalently that H_A is probably true, and that $\mu \neq \mu_0$.

Example 4.1 The stoichiometric factor for percent uranium of UO_2 powder manufactured by a given facility is 87.6%. While a batch of UO_2 powder is being processed, five powder samples are randomly selected from the batch and analyzed for percent uranium. Assume that the percent uranium measured from the individual samples has an approximate normal distribution with mean μ and variance $\sigma^2 = (0.0342)^2$. It is of interest to test the hypothesis

$$H_0: \mu = 87.6$$

against the alternative

$$H_A: \mu = \mu_A \neq 87.6$$

where $\alpha = 0.05$ is the specified probability of making a Type I error. From Table A3, $z_{1-\alpha/2} = z_{0.975} = 1.96$.

The five sample values are 87.627, 87.649, 87.642, 87.571, and 87.637. The sample mean is $\bar{X} = 87.625$. The test statistic is computed from Equation 4.6 as

$$Z = \frac{87.625 - 87.6}{0.0342/\sqrt{5}} = 1.63$$

Because $-1.96 < 1.63 < 1.96$, H_0 is not rejected, and it is concluded that the sample mean percent uranium for this batch of UO_2 powder is not significantly different than the stoichiometric value of $\mu_0 = 87.6\%$. That is, there is insufficient evidence to conclude that the true mean percent uranium for this batch is different than the stoichiometric value of $\mu_0 = 87.6\%$. Thus, the batch is deemed acceptable with respect to its percent uranium.

4.2.1.2 One-Sided Tests

One-sided hypotheses are more specific than two-sided hypotheses because a one-sided alternative states the direction of the difference between μ_0 and μ_A that the investigator wishes to detect. There are two cases to consider, and they are presented here without examples because they are simply modifications of the two-sided test.

It is sometimes of interest to test the hypothesis

$$H_0: \mu \leq \mu_0$$

against the alternative

(4.9)

$$H_A: \mu = \mu_A > \mu_0$$

In this case, a true mean μ which is less than or equal to μ_0 is "acceptable," whereas, if μ is larger than μ_0 , the investigator wishes to detect the difference. In defining the rejection region for this test, observe that a negative value of the test statistic Z would never lead to the conclusion that H_0 is false. Thus, the rejection region is in the upper tail of the $N(0,1)$ distribution. For a specified value of α , H_0 is rejected if

$$Z \geq z_{1-\alpha}$$

(4.10)

Otherwise, H_0 is not rejected.

Similarly, to test the hypothesis

$$H_0: \mu \geq \mu_0$$

against the alternative

(4.11)

$$H_A: \mu = \mu_A < \mu_0$$

the rejection region is in the lower tail of the $N(0,1)$ distribution. Thus, for a specified value of α , H_0 is rejected if

$$Z \leq -z_{1-\alpha} \quad (4.12)$$

Otherwise, H_0 is not rejected.

4.2.1.3 Probability of Type II Error and Power

In testing the hypotheses given by Equations 4.5, 4.9, and 4.11, the probability α of a Type I error, and thus the rejection region for the test, is specified in advance. The probability β of a Type II error, however, depends upon the value of α , the sample size n , the variance σ^2 , and the true value of the mean μ .

Consider the two-sided hypothesis given by Equation 4.5, and suppose that H_A is true. The distribution of \bar{X} is $N(\mu_A, \sigma^2/n)$, and it follows that the distribution of the test statistic is

$$Z \sim N\left(\frac{\mu_A - \mu_0}{\sigma/\sqrt{n}}, 1\right)$$

The value of β is computed by noting that H_0 is not rejected if $-z_{1-\alpha/2} < Z < z_{1-\alpha/2}$. A case where $\mu_A > \mu_0$ is illustrated in Fig. 4.2. A case where

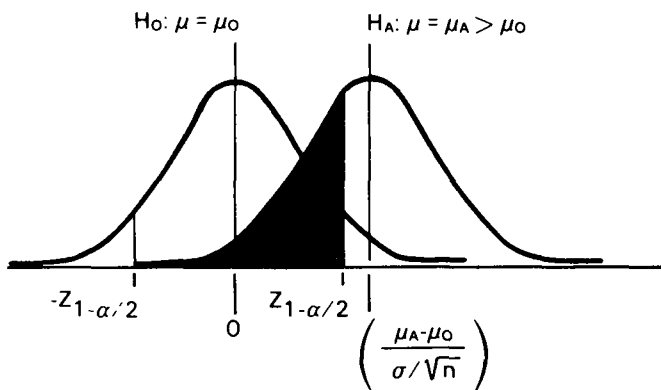


Figure 4.2 Distribution of Z under H_0 and H_A when $\mu_A > \mu_0$.

$\mu_A < \mu_0$ is illustrated in Figure 4.3. In each figure, the probability of not rejecting H_0 is shown as the shaded portion and is computed from the formula

$$\beta = \Phi \left(z_{1-\alpha/2} - \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} \right) - \Phi \left(-z_{1-\alpha/2} - \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} \right) \quad (4.13)$$

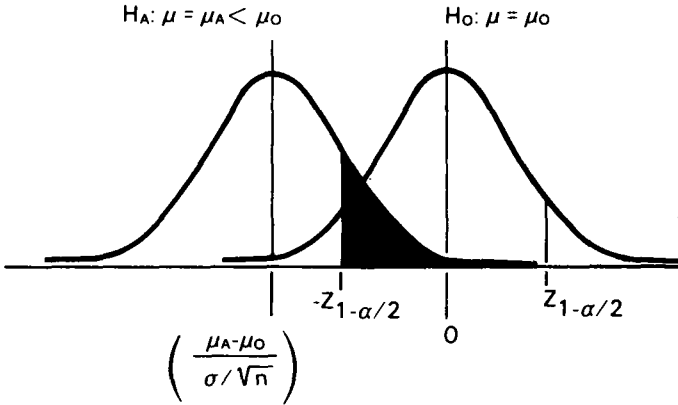


Figure 4.3 Distribution of Z under H_0 and H_A when $\mu_A < \mu_0$.

For the one-sided hypothesis given by Equation 4.9, β is the probability that $Z < z_{1-\alpha}$ when $\mu = \mu_A > \mu_0$. This is given by

$$\beta = \Phi \left(z_{1-\alpha} - \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} \right) \quad (4.14)$$

For the one-sided hypothesis given by Equation 4.11, β is the probability that $Z > -z_{1-\alpha}$ when $\mu = \mu_A < \mu_0$. This is given by

$$\beta = 1 - \Phi \left(-z_{1-\alpha} - \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} \right) \quad (4.15)$$

For a sampling plan, an experiment, or a testing application that is being planned, a recommended practice is to evaluate β for a range of values of μ_A and then plot either β or $1 - \beta$ vs. μ_A . The plot of β vs. μ_A is called an **operating characteristic curve**, or **OC curve**. The plot of $1 - \beta$ vs. μ_A is called a **power curve**. Either type of curve is useful for determining which values of μ_A are likely to result in rejecting H_0 . For a hypothesis test to be informative and worthwhile, the probability of rejecting H_0 (i.e., the power of the test) should be high when the difference between μ_0 and μ_A is large enough to be of practical importance to the investigator.

Example 4.2 For the sampling plan and hypothesis test described in Example 4.1, determine β for values of μ_A from 87.56 to 87.64 in increments of 0.01. Then plot both the OC curve and the power curve for this range of μ_A values.

To illustrate the procedure for determining β , consider $\mu_A = 87.56$. First, compute the quantity

$$\frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} = \frac{87.56 - 87.6}{0.0342/\sqrt{5}} = -2.62$$

and note that $z_{1-\alpha/2} = 1.96$. Substituting these values into Equation 4.13 gives

$$\begin{aligned}\beta &= \Phi(1.96 + 2.62) - \Phi(-1.96 + 2.62) = \Phi(4.58) - \Phi(0.66) \\ &= 1.0 - 0.74537 = 0.25463\end{aligned}$$

Similarly, when $\mu_A = 87.58$,

$$\frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} = -1.31$$

and

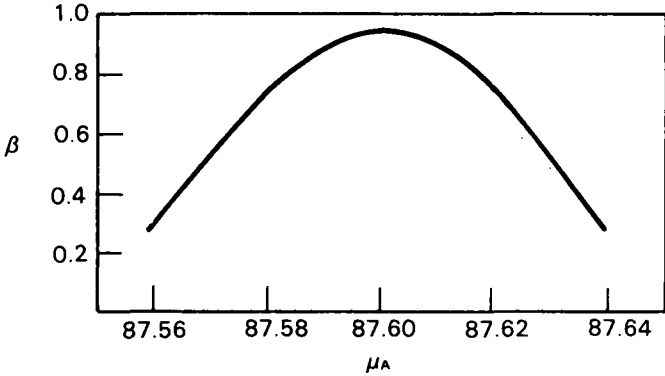
$$\beta = \Phi(3.27) - \Phi(-0.65) = 0.99946 - 0.25785 = 0.74161$$

Continuing in this manner and rounding the probabilities to two significant digits, the following table is constructed:

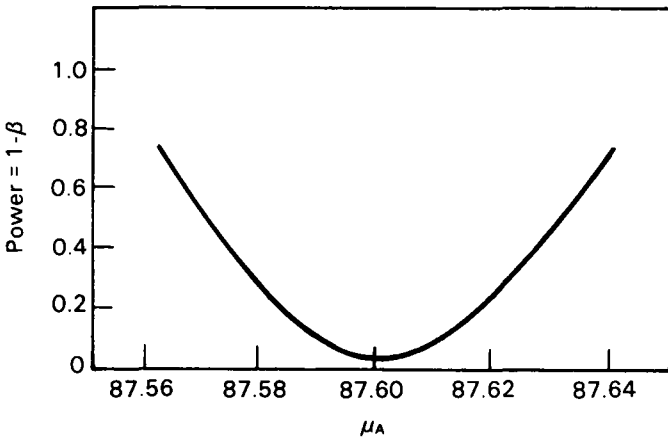
μ_A	β	$1 - \beta$
87.56	0.25	0.75
87.57	0.50	0.50
87.58	0.74	0.26
87.59	0.90	0.10
87.60	0.95	0.05
87.61	0.90	0.10
87.62	0.74	0.26
87.63	0.50	0.50
87.64	0.25	0.75

Note that β decreases ($1 - \beta$ increases) as μ_A gets farther away in either direction from $\mu_0 = 87.6$. Also note that when $\mu_A = \mu_0$ (i.e., when H_0 is true), $\beta = 1 - \alpha = 0.95$ and $1 - \beta = \alpha = 0.05$. Of course, $\alpha = 0.05$ is simply the prespecified probability of rejecting H_0 when H_0 is true.

The OC curve for this test is constructed by plotting β vs. μ_A and is shown below.



The power curve for this test is constructed by plotting $1 - \beta$ vs. μ_A and is shown below.



Interpretation of either of these curves depends upon which values of μ_A would cause concern (or are unacceptable). Suppose that a mean percent uranium value of 87.57 is unacceptably low. Notice from the OC curve and power curve that $\beta = 1 - \beta = 0.50$ when $\mu_A = 87.57$. That is,

when $\mu_A = 87.57$, there is a 50% chance of not rejecting H_0 and a 50% chance of rejecting H_0 . This indicates that the sampling plan and hypothesis testing procedure described in Example 4.1 are inadequate for detecting when the mean percent uranium is 87.57. Example 4.3 illustrates that one way to resolve this is to take a larger sample.

4.2.1.4 Sample Size

For the two-sided hypothesis given by Equation 4.5, the investigator might state that if the absolute difference $|\mu_A - \mu_0|$ is greater than or equal to a specified value, then the power of the test must be greater than or equal to a specified level. Suppose that in the application described in Example 4.1, it is required that if the true mean percent uranium is smaller than or equal to 87.57 or larger than or equal to 87.63, then the power of the test must be at least 0.90. That is, if $|\mu_A - 87.6| \geq 0.03$, then $1 - \beta \geq 0.90$. The results of Example 4.2 indicate that $1 - \beta = 0.50$ when $|\mu_A - 87.6| = 0.03$. This is clearly not in accordance with the requirements.

The results of Example 4.2 indicate that the power of the two-sided hypothesis test increases (β decreases) as μ_A gets farther away in either direction from μ_0 . In fact, an examination of Equation 4.13 reveals that β decreases (the power of the test increases) as the quantity $(\mu_A - \mu_0)/(\sigma/\sqrt{n})$ gets farther away from zero. In planning a hypothesis test, a value of $(\mu_A - \mu_0)$ is specified that is of practical importance to the experimenter. Values of α and either β or $1 - \beta$ are also specified. That is, all quantities in Equation 4.13 are specified except σ/\sqrt{n} , the standard deviation of \bar{X} . Thus, a value of σ/\sqrt{n} must be determined such that Equation 4.13 holds when values of α , β , and $(\mu_A - \mu_0)$ are specified.

Because σ is assumed to be known, a common practice is to determine the sample size n required to satisfy Equation 4.13 when all other quantities are specified. This problem can be simplified somewhat for applications where the specified values of α and β are small (for example, <0.25). Note in Fig. 4.2 that when $\mu_A > \mu_0$, especially if $(\mu_A - \mu_0)/(\sigma/\sqrt{n}) > z_{1-\alpha/2}$, a negligible portion of the shaded region extends below $z_{1-\alpha/2}$. In such cases the second term on the right-hand side of Equation 4.13 is practically zero. Then, from the symmetry of the normal pdf, $z_\beta = -z_{1-\beta}$, so that

$$\beta = \Phi(z_{1-\beta}) \cong \Phi[z_{1-\alpha/2} - \sqrt{n}(\mu_A - \mu_0)/\sigma]$$

or equivalently

$$-z_{1-\beta} \cong z_{1-\alpha/2} - \sqrt{n}(\mu_A - \mu_0) / \sigma$$

Rearranging terms and squaring gives

$$n \cong \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2} \quad (4.16)$$

Equation 4.16 can also be derived from a similar argument using the case illustrated in Fig. 4.3 where $\mu_A < \mu_0$. Thus, for specified values of σ^2 , α , β , and $(\mu_A - \mu_0)$, Equation 4.16 gives a good approximation of the required sample size for a two-sided test. Because sample sizes are integers, a noninteger value computed from Equation 4.16 is rounded up to the next integer.

The sample size formula which applies to either of the one-sided tests is given by

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2} \quad (4.17)$$

which is Equation 4.16 with $z_{1-\alpha}$ in place of $z_{1-\alpha/2}$. Equation 4.17, however, is not an approximation.

In many applications, the costs and resource requirements associated with sampling are a major consideration. Thus, using Equation 4.16 or 4.17 alone might not be wholly satisfactory, because the result might indicate that an infeasibly large sample size is required to achieve a specified level of power. In such cases, there are other factors to consider which sometimes can be altered to achieve both an affordable sampling plan and a hypothesis test with satisfactory properties.

Recall from the previous discussion that for a fixed value of α , the power of a hypothesis test increases (β decreases) as $(\mu_A - \mu_0)/(\sigma/\sqrt{n})$ gets farther away from zero in a specified direction, depending upon which hypothesis is being tested. Thus, when α , $(\mu_A - \mu_0)$, and σ are fixed, increasing the sample size n pushes $(\mu_A - \mu_0)/(\sigma/\sqrt{n})$ farther away from zero, which increases the power of the test. If there is an upper limit on the sample size, however, a desired level of power can still be achieved if σ can be decreased enough to satisfy Equation 4.13, 4.14, or 4.15 when α , β , $(\mu_A - \mu_0)$, and n are fixed. In practice, this requires careful examination of the process or population being sampled and the methods of sampling, sample analysis, measurement, and so on with the objective of identifying sources of variation that can either be better controlled or eliminated.

It is not always possible or practical to reduce σ , and even when σ can be reduced, the amount of reduction possible may not be enough to

achieve the level of power desired. In such cases, the specified values of α , $(\mu_A - \mu_0)$, and $(1 - \beta)$ or β should be reviewed. Perhaps the desired level of power is unrealistic or excessively strict when the purpose of the hypothesis test and some of the practical limitations are considered. If so, a less restrictive set of specified values should be considered. If acceptable, the result could be an affordable sampling plan and a hypothesis test that has satisfactory properties with respect to the particular application.

When compromises are made between a desired level of power and one that is practical and affordable, it is essential that all parties involved be aware of the properties and limitations of the resulting hypothesis test. A power curve or OC curve is invaluable for this purpose because it illustrates in a concise form the probability of rejecting (or not rejecting) H_0 for any value of μ_A .

Example 4.3 How can the sampling plan described in Example 4.1 be modified to ensure that $1 - \beta \geq 0.90$ if $\mu \leq 87.57$ or $\mu \geq 87.63$?

First, note that

$$(87.57 - 87.6)^2 = (87.63 - 87.6)^2 = (0.03)^2$$

$$z_{1-\beta} = z_{0.90} = 1.28$$

$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

$$\sigma^2 = (0.0342)^2$$

Substituting these values into Equation 4.16 gives

$$n \cong \frac{(1.96 + 1.28)^2 (0.0342)^2}{(0.03)^2} = 13.64 \cong 14$$

That is, because the sample size must be an integer, n is always rounded up to be conservative.

The new sampling plan is to randomly select 14 powder samples from a batch. The percent uranium will be determined for each powder sample, and the test statistic

$$Z = \frac{\bar{X} - 87.6}{\sigma/\sqrt{14}}$$

will be computed. If $Z \leq -1.96$ or $Z \geq 1.96$, reject H_0 .

A sample of size 14 is too costly for this particular application. Suppose, however, that it is possible to refine the method of analyzing the powder samples such that σ is reduced from 0.0342 to 0.03. Suppose also that upon reviewing the hypothesis testing procedure, it is decided that a two-sided test is not really necessary because there is a natural upper limit on the percent uranium (i.e., stoichiometric UO_2). Thus, a sample mean significantly larger than 87.6% would indicate a bias in the method of analyzing the powder samples rather than an undesirable percent uranium factor. Because the analysis procedure is monitored and maintained independently of this application, the upper rejection region, $Z > z_{1-\alpha/2}$, is of little practical interest. A sample mean significantly less than 87.6%, however, could indicate that impurities are present or that the composition of the powder is not correct.

Thus, suppose that the scope of the problem is changed so that the hypothesis to be tested is

$$H_0: \mu \geq 87.6$$

against the alternative

$$H_A: \mu < 87.6$$

and that the significance level is still $\alpha = 0.05$. The power requirements have been relaxed slightly, such that when $\mu \leq 87.565$ the probability of rejecting H_0 should be $1 - \beta = 0.90$. The required sample size for this modified plan is determined by substituting the values

$$(\mu_A - \mu_0) = -0.035$$

$$z_{1-\beta} = 1.28$$

$$z_{1-\alpha} = 1.65$$

$$\sigma^2 = (0.03)^2$$

into Equation 4.17, which gives

$$n = \frac{(1.65 + 1.28)^2 (0.03)^2}{(-0.035)^2} = 6.31 \cong 7$$

Thus, the modified plan is to randomly select seven powder samples from a batch. The percent uranium will be determined for each sample, and the test statistic

$$Z = \frac{\bar{X} - 87.6}{0.03/\sqrt{7}}$$

will be computed. H_0 will be rejected if $Z \leq -1.65$.

The points for constructing a power curve or OC curve for this new plan are determined from Equation 4.15 and summarized in the following table.

μ_A	β	$1 - \beta$
87.550	0.003	0.997
87.555	0.010	0.990
87.560	0.030	0.970
87.565	0.076	0.924
87.570	0.159	0.841
87.575	0.288	0.712
87.580	0.456	0.544
87.590	0.776	0.224
87.600	0.950	0.050
87.700	0.994	0.006

Note that when $\mu_A = 87.565\%$, the power is $1 - \beta = 0.924$ which exceeds the specified value of 0.90. This is because the sample size $n = 7$ is larger than the value $n = 6.31$ computed from Equation 4.17. This illustrates that rounding n to the next larger integer is conservative.

This example illustrates how a review of the purpose of a hypothesis test can result in redefining both the hypothesis and the power requirements. This, together with a reduction in σ through better control of an analysis procedure, has resulted in a required sample size which is half as large as that required to meet the previous specifications.

4.2.1.5 Confidence Intervals and Hypothesis Testing

There is a close relationship between hypothesis testing and confidence intervals. Consider testing the two-sided hypothesis given by Equation 4.5. Note from Figure 4.1 and Equation 4.6, that if $\mu = \mu_0$, the expression $-z_{1-\alpha/2} < Z < z_{1-\alpha/2}$ is equivalent to

$$-z_{1-\alpha/2} < \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha/2}$$

or

$$-z_{1-\alpha/2}(\sigma/\sqrt{n}) < \bar{X} - \mu_0 < z_{1-\alpha/2}(\sigma/\sqrt{n})$$

This can be rearranged to the form

$$\bar{X} - z_{1-\alpha/2}(\sigma/\sqrt{n}) < \mu_0 < \bar{X} + z_{1-\alpha/2}(\sigma/\sqrt{n})$$

Thus, the two-sided hypothesis test is equivalent to the following procedure: Construct a $(1 - \alpha)100\%$ confidence interval for μ (Section 3.5.2). If μ_0 is inside the confidence interval, do not reject H_0 . If μ_0 is outside the confidence interval, reject H_0 .

Example 4.4 Rework Example 4.1 using the confidence interval approach. Recall that $\mu_0 = 87.60$, $\bar{X} = 87.625$, $\alpha = 0.05$, $z_{1-\alpha/2} = 1.96$, $\sigma = 0.0342$, and $n = 5$. Thus, a 95% confidence interval for μ is given by

$$87.625 \pm 1.96(0.0342)/\sqrt{5}$$

or

$$(87.595, 87.655)$$

Because $\mu_0 = 87.60$ is inside this interval, H_0 is not rejected. This, of course, agrees with the conclusion reached in Example 4.1.

4.2.2 Tests About the Mean, Variance Unknown

We now consider the case where a random variable X has a normal distribution with mean μ and variance σ^2 , both unknown, and we wish to test hypotheses about μ . Although the normality assumption is required in theory, slight departures from normality are not serious in practice, particularly if the sample size is not small.

Assume that a random sample of size n is drawn and that \bar{X} and S^2 are the sample mean and sample variance computed from the sample data using Equations 3.11 and 3.13.

The hypothesis to be tested is

$$H_0: \mu = \mu_0$$

against the alternative

$$H_A: \mu = \mu_A \neq \mu_0$$

with a specified probability α of a Type I error.

Since σ^2 is unknown, substitute the sample variance S^2 for σ^2 in Equation 4.6 to compute the test statistic t , given by

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (4.18)$$

If H_0 is true, t has a Student's t -distribution with $n - 1$ degrees of freedom (Section 3.5.2). Thus, H_0 is rejected if either

$$t \leq -t_{1-\alpha/2}(n-1)$$

or

$$t \geq t_{1-\alpha/2}(n-1)$$

where $t_{1-\alpha/2}(n-1)$ is obtained from Table A5. Hypothesis tests that are based upon the Student's t -distribution are commonly referred to as t -tests.

The one-sided tests are simply modifications of the two-sided test procedure described above. To test the hypothesis

$$H_0: \mu \leq \mu_0$$

against the alternative

$$H_A: \mu = \mu_A > \mu_0$$

at the α level of significance, reject H_0 if $t \geq t_{1-\alpha}(n-1)$.

To test the hypothesis

$$H_0: \mu \geq \mu_0$$

against the alternative

$$H_A: \mu = \mu_A < \mu_0$$

at the α level of significance, reject H_0 if $t \leq -t_{1-\alpha}(n - 1)$.

Example 4.5 In Example 4.3, a sampling plan and hypothesis test are developed for monitoring the percent uranium in batches of UO_2 powder. In that development, there is an underlying assumption that σ can be reduced from 0.0342 to 0.03. Suppose that the necessary steps have been taken to reduce σ , but their effect on σ has not yet been established. That is, σ is unknown. A new batch of UO_2 powder is ready to be tested, however, and a decision is made to proceed by randomly selecting $n = 7$ powder samples and testing the hypothesis

$$H_0: \mu \geq 87.60$$

against the alternative

$$H_A: \mu = \mu_A < 87.60$$

at the $\alpha = 0.05$ significance level using the t-test described in this section.

The percent uranium values for the seven powder samples are:

$$87.559, 87.552, 87.587, 87.600, 87.517, 87.609, 87.565$$

The sample mean, sample standard deviation, and test statistic are computed as

$$\bar{X} = 87.5699$$

$$S = 0.0316$$

and

$$t = \frac{87.5699 - 87.60}{0.0316/\sqrt{7}} = -2.52$$

From Table A5, $t_{0.95}(6) = 1.94$. Because $-2.52 < -1.94$, H_0 is rejected. It is concluded that the true mean percent uranium for this batch of UO_2 powder is less than 87.60%.

When H_0 is false, the test statistic t given by Equation 4.18 follows a distribution known as the **noncentral t-distribution** with $n - 1$ degrees of freedom and noncentrality parameter $\lambda = \sqrt{n}(\mu_A - \mu_0)/\sigma$. Thus, β , the probability of a Type II error for a specified value of μ_A , is determined from the noncentral t-distribution. A detailed presentation of the noncentral t-distribution is beyond the scope of this book. Tables and charts are provided by Ferris, Grubbs, and Weaver (1946) and Bowker and Lieberman (1972). Use of these tables and charts, however, requires values of λ which involve the unknown parameter σ . Thus, results of implementing the noncentral t-distribution are approximate.

For most situations in practice, Equation 4.13, 4.14, or 4.15 can be used to approximate β for a specified value of μ_A . For a given value of σ , this approximation underestimates β , especially if $n < 30$, but it provides valuable information to the practitioner. Another possible drawback is that Equations 4.13, 4.14, and 4.15 require specifying a value for σ . There are a number of possibilities for selecting a value of σ . If a similar experiment or a "pilot study" has been completed, the sample variance S^2 could be used in place of σ^2 in the approximation. If no experimental data are available, then an estimate of σ must come from prior studies, scientific laws, or guesswork. Another possibility is to express $(\mu_A - \mu_0)$ as a multiple of σ , so that the OC curve is β plotted against values of $(\mu_A - \mu_0)/\sigma$.

When σ is unknown, an equation for determining a required sample size can be derived from the noncentral t-distribution. Because the degrees of freedom depend upon the sample size, however, the solution to this equation is iterative and involves numerical analysis techniques. To avoid this difficulty, Guenther (1981) proposes an approximate solution which is a modification of Equation 4.16 or Equation 4.17. For testing the two-sided hypothesis at the α level of significance with power $1 - \beta$ for a specified value of $(\mu_A - \mu_0)$, the required sample size is approximately

$$n \cong \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2} + 0.5 (z_{1-\alpha/2})^2 \quad (4.19)$$

where σ^2 is replaced by the best available estimate. Since the sample size must be an integer, n is always rounded up to be conservative. The approximate required sample size for either of the one-sided tests is computed by substituting $z_{1-\alpha}$ in place of $z_{1-\alpha/2}$ in Equation 4.19. Guenther (1981) states that this approximation is quite accurate and usually gives the same result that would be obtained by iterative numerical methods.

Example 4.6 Continuing with the UO_2 powder example, suppose that until the true value of σ can be established, the t-test will be used to monitor the percent uranium in batches of UO_2 powder. There is some concern, however, that the sample size $n = 7$ may be inadequate to ensure that $1 - \beta \geq 0.90$ when $(\mu_A - \mu_0) \leq -0.035$. Should the sample size be changed for future testing?

Recall from Examples 4.4 and 4.5 that for each batch of UO_2 powder, the hypothesis

$$H_0: \mu \geq 87.6$$

vs. the alternative

$$H_A: \mu = \mu_A < 87.6$$

is to be tested at the $\alpha = 0.05$ level of significance. Thus, using the value $S = 0.0316$ from Example 4.5 as an estimate of σ and noting that

$$\begin{aligned}(\mu_A - \mu_0)^2 &= (-0.035)^2 \\ z_{1-\beta} &= 1.28\end{aligned}$$

and

$$z_{1-\alpha} = 1.65$$

the approximate required sample size, computed from Equation 4.19, is

$$n \cong \frac{(1.65 + 1.28)^2 (0.0316)^2}{(-0.035)^2} + 0.5 (1.65)^2 = 8.36 \cong 9$$

That is, the sample size should be increased to $n = 9$ until the value of σ

is established, at which time Equation 4.17 should be used to determine the sample size for future testing.

4.2.3 Tests About the Variance

There are situations in practice where it is useful to test the null hypothesis that the variance σ^2 of a normal distribution is equal to a specified value, say σ_0^2 . This arises, for example, when a quality standard pertaining to the amount of variability allowable in some measurement specifies a value $\sigma^2 = \sigma_0^2$. It also arises when σ^2 is predicted from some theory that is to be tested, or when σ^2 is known for a population to which sample data are being compared.

Unlike the tests about the mean, tests about the variance are quite sensitive to departures from the normality assumption. Thus, we must assume rather strictly that $X \sim N(\mu, \sigma^2)$, where μ and σ^2 are unknown, and that X_1, X_2, \dots, X_n is a random sample of n observations on X . When the normality assumption is in doubt, nonparametric methods should be used to test hypotheses about the variance (Sections 9.2.2 and 9.3.2).

Suppose that we wish to test the hypothesis

$$H_0: \sigma^2 = \sigma_0^2$$

against the alternative (4.20)

$$H_A: \sigma^2 = \sigma_A^2 \neq \sigma_0^2$$

at the α level of significance. The test statistic is given by

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \tag{4.21}$$

where S^2 is the sample variance computed from the n observations using Equation 3.13. When H_0 is true, χ^2 has a chi-square distribution with $n-1$ degrees of freedom. Thus, H_0 is rejected if either

$$\chi^2 \leq \chi_{\alpha/2}^2 (n-1)$$

or (4.22)

$$\chi^2 \geq \chi_{1-\alpha/2}^2(n-1)$$

where $\chi_{\alpha/2}^2(n-1)$ and $\chi_{1-\alpha/2}^2(n-1)$ are obtained from Table A4.

The one-sided tests are more common, particularly the test of the hypothesis

$$H_0: \sigma^2 \leq \sigma_0^2$$

against the alternative

(4.23)

$$H_A: \sigma^2 = \sigma_A^2 > \sigma_0^2$$

H_0 is rejected at the α level of significance if

$$\chi^2 \geq \chi_{1-\alpha}^2(n-1) \quad (4.24)$$

The less common one-sided hypothesis is

$$H_0: \sigma^2 \geq \sigma_0^2$$

against the alternative

(4.25)

$$H_A: \sigma^2 = \sigma_A^2 < \sigma_0^2$$

Here H_0 is rejected at the α level of significance if

$$\chi^2 \leq \chi_{\alpha}^2(n-1) \quad (4.26)$$

Example 4.7 In an automated process, UO_2 powder containers are filled to a prescribed weight, sealed, and labeled. Periodically, a random sample of six filled containers is weighed to monitor the variability in the automatic filling process. The hypothesis of interest is

$$H_0: \sigma^2 \leq (0.20 \text{ kg})^2$$

against the alternative

$$H_A: \sigma^2 > (0.20 \text{ kg})^2$$

and the level of significance is $\alpha = 0.10$.

The most recent sample of six containers had the following weights in kg: 22.06, 22.52, 22.34, 22.90, 22.86, 22.18. From these weights, the value $S^2 = (0.3488)^2$ is computed. The test statistic is

$$\chi^2 = \frac{(5)(0.3488)^2}{(0.20)^2} = 15.21$$

From Table A4, $\chi_{0.90}^2(5) = 9.24$. Since $\chi^2 = 15.21 > 9.24$, reject H_0 . Conclude that the variance in filled container weights exceeds $(0.20 \text{ kg})^2$.

Operating characteristic curves for hypothesis tests about σ^2 are given by Ferris, Grubbs, and Weaver (1946) and Bowker and Lieberman (1972). They provide charts where β is plotted against σ_A/σ_0 for various sample sizes. To compute β or $1 - \beta$ for a specified value of σ_A/σ_0 is not difficult but can involve interpolation because of the usually limited tables available for the chi-square distribution. The procedure for computing β is illustrated for the one-sided hypothesis given by Equation 4.23.

For given values of n , α , σ_0^2 , and σ_A^2 , the rejection region of the test includes values of $\chi^2 \geq \chi_{1-\alpha}^2(n-1)$. That is, under H_0 , χ^2 has a chi-square distribution with $n - 1$ degrees of freedom. When H_A is true, the probability of accepting H_0 is given by

$$\beta = \Pr \left[\chi^2 < \frac{\sigma_0^2}{\sigma_A^2} \chi_{1-\alpha}^2(n-1) \right] \quad (4.27)$$

The value of β is determined as follows: Obtain $\chi_{1-\alpha}^2(n-1)$ from Table A4, and multiply by (σ_0^2/σ_A^2) . Enter Table A4 and locate the row with $n-1$ degrees of freedom in the left margin. On this row, locate the computed value $(\sigma_0^2/\sigma_A^2) \chi_{1-\alpha}^2(n-1)$, and read β from the column heading. The power of the test at this value of σ_A^2 is $1 - \beta$.

Example 4.8 For the problem described in Example 4.7, find β when $\sigma^2 = \sigma_A^2 = (0.25)^2$. Recall that $\chi_{0.90}^2(5) = 9.24$, $n = 6$, and $\sigma_0^2 = (0.20)^2$. Then

$$\frac{\sigma_0^2}{\sigma_A^2} \chi_{0.90}^2(5) = \frac{(0.20)^2}{(0.25)^2} (9.24) = 5.91$$

Entering Table A4 with $n - 1 = 5$ degrees of freedom, we do not find the value 5.91, but we see that the value 5.13 corresponds to $\beta = 0.60$ and that 6.06 corresponds to $\beta = 0.70$.

Using linear interpolation, an approximate value of β for 5.91 is

$$\beta \cong 0.70 - (0.70 - 0.60) \frac{(6.06 - 5.91)}{(6.06 - 5.13)} = 0.70 - 0.016 = 0.684$$

That is, when $\sigma^2 = (0.25)^2$, $\beta = 0.684$, and the power of the test is $1 - \beta = 0.316$.

A sample size can be determined which ensures that the power of the test $1 - \beta$ is at a specified level for a given value of σ_A^2/σ_0^2 . The procedure requires specified values of α , β , and $\lambda = (\sigma_A^2/\sigma_0^2)$. For the one-sided hypothesis given by Equation 4.23, enter Table A4, compare the columns headed by β and $1 - \alpha$, and locate the row where $\lambda\chi_{\beta}^2 = \chi_{1-\alpha}^2$. This row has $n - 1$ degrees of freedom in the left margin, and the required sample size is $n = (n - 1) + 1$ (i.e., the degrees of freedom plus one).

Example 4.9 For the problem described in Example 4.7, determine the sample size required to ensure that $1 - \beta = 0.80$ when $(\sigma_A^2/\sigma_0^2) = 2.0$.

Enter Table A4 and locate the columns headed by 0.20 and 0.90. Note that when $n - 1 = 18$, $2\chi_{0.20}^2(18) = 25.8$ and $\chi_{0.90}^2(18) = 26.0$. Also, when $n - 1 = 19$, $2\chi_{0.20}^2(19) = 27.4$ and $\chi_{0.90}^2(19) = 27.2$. To be conservative, choose the larger degrees of freedom value $n - 1 = 19$, which gives the required sample size $n = 19 + 1 = 20$.

Thus, if it is desirable to have $1 - \beta = 0.80$ when $\sigma^2 = (2)(0.20)^2$, the sample size must be increased to $n = 20$.

4.3 INDEPENDENT SAMPLES FROM TWO NORMAL DISTRIBUTIONS

In this section, hypothesis testing procedures are presented for comparing the parameters of two normal distributions when each of the distributions has been sampled independently of the other. The tests fall under three broad categories: comparing the means when the variances are known, comparing the means when the variances are unknown, and comparing the variances.

The tests for comparing means are based upon the sample means. Thus, although the normality assumption is necessary to develop the testing procedures, the procedures are applicable for comparing the means of many different types of distributions if the sample sizes satisfy the rules of thumb for applying the Central Limit Theorem. The testing procedure for comparing the variances, however, is sensitive to departures from the normality assumption. In either case, if there is any reason to doubt the validity of the normality assumption, one of the nonparametric techniques presented in Chapter 9 should be used to test the hypothesis of interest.

4.3.1 Comparison of Two Means, Variances Known

Suppose that we are interested in comparing the means of two normal distributions, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where μ_1 and μ_2 are unknown, whereas σ_1^2 and σ_2^2 are assumed known. For example, it might be desirable to compare the mean percent uranium in two batches of UO_2 powder which were prepared under different process conditions. Assume that \bar{X}_1 is the sample mean computed from a random sample of size n_1 from $N(\mu_1, \sigma_1^2)$ and that \bar{X}_2 is the sample mean computed from a random sample of size n_2 from $N(\mu_2, \sigma_2^2)$. Assume further that the two random samples are drawn independently of each other.

Consider first testing the two-sided hypothesis

$$H_0: \mu_1 - \mu_2 = \delta_0$$

against the alternative

(4.28)

$$H_A: \mu_1 - \mu_2 = \delta_A \neq \delta_0$$

at the α level of significance. Here, δ_0 is a specified difference between μ_1 and μ_2 , and δ_A is any alternative value of the difference. For many applications δ_0 is zero.

The testing procedure is based upon $\bar{X}_1 - \bar{X}_2$, the difference in the sample means. Using the results presented in Section 2.7.1, it can be shown that

$$(\bar{X}_1 - \bar{X}_2) \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

It follows that if the null hypothesis is true, then the test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{1/2}} \quad (4.29)$$

has the $N(0,1)$ distribution. Thus, H_0 is rejected if either

$$Z \leq -z_{1-\alpha/2}$$

or

$$Z \geq z_{1-\alpha/2}$$

(4.30)

Example 4.10 A facility which produces UO_2 pellets is interested in comparing the mean percent uranium of two batches of pellets produced under different operating conditions. It has been established from experience that the percent uranium of pellets in a batch is normally distributed with unknown mean μ and with variance $\sigma^2 = 0.0055$. The variance is due mostly to the method of analysis and remains practically constant from batch to batch. A random sample of $n_1 = 8$ pellets is taken from the first batch, and a random sample of $n_2 = 12$ pellets is taken from the second batch. The data are:

Batch 1	batch 2
88.056	87.939
88.088	87.883
88.044	88.005
88.015	88.064
87.897	88.001
88.039	87.977
87.950	87.881
88.113	87.946
	88.107
	87.970
	87.923
	88.119

The hypothesis to be tested is

$$H_0: \mu_1 - \mu_2 = 0$$

against the alternative

$$H_A: \mu_1 - \mu_2 = \delta_A \neq 0$$

where μ_1 and μ_2 are the true mean percent uranium for batches 1 and 2, respectively. A significance level $\alpha = 0.10$ is designated for this test.

The sample means are $\bar{X}_1 = 88.0253$ and $\bar{X}_2 = 87.9846$, with $n_1 = 8$ and $n_2 = 12$. Substituting these values and $\sigma_1^2 = \sigma_2^2 = 0.0055$ into Equation 4.29, the value of the test statistic is

$$Z = \frac{(88.0253 - 87.9846) - 0}{\left(\frac{0.0055}{8} + \frac{0.0055}{12} \right)^{1/2}} = 1.20$$

From Table A3, $z_{1-\alpha/2} = z_{0.95} = 1.65$. Because $-1.65 < 1.20 < 1.65$, H_0 is not rejected. There is not sufficient evidence to conclude that the two batches of pellets differ in mean percent uranium.

The one-sided tests are modifications of the two-sided test. To test the hypothesis

$$H_0: \mu_1 - \mu_2 \leq \delta_0$$

against the alternative

$$(4.31)$$

$$H_A: \mu_1 - \mu_2 = \delta_A > \delta_0$$

at the α level of significance, reject H_0 if

$$Z \geq z_{1-\alpha} \quad (4.32)$$

To test the hypothesis

$$H_0: \mu_1 - \mu_2 \geq \delta_0$$

against the alternative

(4.33)

$$H_A: \mu_1 - \mu_2 = \delta_A < \delta_0$$

at the α level of significance, reject H_0 if

$$Z \leq -z_{1-\alpha} \quad (4.34)$$

Power curves and OC curves are constructed for tests about $\mu_1 - \mu_2$ by analogy with the procedure given in Section 4.2.1.3. For example, consider the two-sided hypothesis given by Equation 4.28. When H_A is true, the probability of making a Type II error is

$$\begin{aligned} \beta = \Phi \left[z_{1-\alpha/2} - (\delta_A - \delta_0) / (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2} \right] \\ - \Phi \left[-z_{1-\alpha/2} - (\delta_A - \delta_0) / (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2} \right] \quad (4.35) \end{aligned}$$

The power curve is constructed by plotting $1-\beta$ over a given range of values of δ_A . Power curves for the one-sided tests are constructed by noting that if $\alpha/2$ is replaced by α in Equation 4.35, the first term on the right-hand side gives β for the hypothesis in Equation 4.31, whereas one minus the second term gives β for the hypothesis in Equation 4.33.

It is possible to derive formulas to compute the approximate sample sizes required to achieve a specified value of β for given values of δ_0 , δ_A , σ_1^2 , σ_2^2 , and α . To use these formulas, however, it is assumed that independent samples of equal size are to be drawn from the two distributions of interest; that is, $n_1 = n_2$. Because the derivation closely follows the single sample case in Section 4.2.1.4, only the formulas are given.

For testing the two-sided hypotheses given by Equation 4.28, the approximate required sample sizes are given by

$$n_1 = n_2 \cong \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\delta_A - \delta_0)^2} \quad (4.36)$$

For testing either of the one-sided hypotheses given by Equations 4.31 and 4.33, the required sample sizes are given by

$$n_1 = n_2 = \frac{(z_{1-\alpha} + z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\delta_A - \delta_0)^2} \quad (4.37)$$

Example 4.11 For the hypothesis test described in Example 4.10, how large would n_1 and n_2 have to be to ensure a 90% chance of rejecting H_0 when there is a difference $|\mu_1 - \mu_2| \geq 0.08$.

For this problem, $\beta = 0.10$, and Table A3 gives $z_{1-\beta} = 1.28$. Because in Example 4.10, $\delta_0 = 0$, it follows that $|\delta_A - \delta_0| = 0.08$. Also, recall that $\alpha = 0.10$, $z_{1-\alpha/2} = 1.65$, and $\sigma_1^2 = \sigma_2^2 = 0.0055$. Substituting these values into Equation 4.36 gives

$$n_1 = n_2 \cong \frac{(1.65 + 1.28)^2 (0.0055 + 0.0055)}{(0.08)^2} = 14.75 \cong 15$$

Thus, a sample of size 15 taken from each batch will ensure a 90% chance of rejecting H_0 when the mean percent uranium in two batches of pellets differs by 0.08 (note that this is 0.08% absolute).

4.3.2 Comparison of Two Means, Variances Unknown

When comparing the means of two normal distributions with unknown variances, there are two cases to consider. If it can be assumed that $\sigma_1^2 = \sigma_2^2$, an exact test is possible. If $\sigma_1^2 \neq \sigma_2^2$, however, an approximate test must be made. In this section, hypothesis testing procedures are given for both cases.

4.3.2.1 Comparison of Two Means, Assuming Equal Variances

Assume that a random variable X_1 has the $N(\mu_1, \sigma_1^2)$ distribution and that a random variable X_2 has the $N(\mu_2, \sigma_2^2)$ distribution, where μ_1 , μ_2 , σ_1^2 , and σ_2^2 are unknown, but $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Consider testing the hypothesis

$$H_0: \mu_1 - \mu_2 = \delta_0$$

against the alternative

$$(4.38)$$

$$H_A: \mu_1 - \mu_2 = \delta_A \neq \delta_0$$

at the α level of significance.

Assume that independent random samples of size n_1 and n_2 are taken on X_1 and X_2 , respectively, and that the sample means and sample variances, \bar{X}_1 , \bar{X}_2 , S_1^2 , and S_2^2 are computed from the sample data. Because both S_1^2 and S_2^2 provide estimates of the common variance, they are com-

bined or pooled to form a single estimator, which is given by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (4.39)$$

This pooled estimator is introduced and discussed in Section 3.5.2.

To test the hypothesis in Equation 4.38, compute the test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{S_p (1/n_1 + 1/n_2)^{1/2}} \quad (4.40)$$

If H_0 is true, t has a Student's t -distribution with $n_1 + n_2 - 2$ degrees of freedom. Thus, reject H_0 if either

$$t \leq -t_{1-\alpha/2}(n_1 + n_2 - 2)$$

or

(4.41)

$$t \geq t_{1-\alpha/2}(n_1 + n_2 - 2)$$

The one-sided hypotheses are treated similarly. To test the hypothesis

$$H_0: \mu_1 - \mu_2 \leq \delta_0$$

against the alternative

(4.42)

$$H_A: \mu_1 - \mu_2 = \delta_A > \delta_0$$

at the α level of significance, reject H_0 if

$$t \geq t_{1-\alpha}(n_1 + n_2 - 2) \quad (4.43)$$

To test the hypothesis

$$H_0: \mu_1 - \mu_2 \geq \delta_0$$

against the alternative

$$(4.44)$$

$$H_A: \mu_1 - \mu_2 = \delta_A < \delta_0$$

at the α level of significance, reject H_0 if

$$t \leq -t_{1-\alpha}(n_1 + n_2 - 2) \quad (4.45)$$

Example 4.12 Rework Example 4.10, assuming that $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and that the value of σ^2 is unknown.

From the sample data in Example 4.10, $n_1 = 8$, $n_2 = 12$, $\bar{X}_1 = 88.0253$, $S_1^2 = 0.005055$, $\bar{X}_2 = 87.9846$, and $S_2^2 = 0.006241$. The pooled estimate of σ^2 is computed from Equation 4.39 as

$$S_p^2 = \frac{7(0.005055) + 11(0.006241)}{18} = 0.005780$$

and the pooled standard deviation is

$$S_p = \sqrt{S_p^2} = 0.0760$$

The test statistic is computed from Equation 4.40 as

$$t = \frac{(88.0253 - 87.9846) - 0}{0.0760 [(1/8) + (1/12)]^{1/2}} = 1.17$$

From Table A5, $t_{0.95}(18) = 1.73$. Because $-1.73 < 1.17 < 1.73$, H_0 is not rejected. There is not sufficient evidence to conclude that the mean percent uranium differs for the two batches of pellets.

When H_0 is false, the test statistic t , given by Equation 4.40, follows the **noncentral t-distribution** with $n_1 + n_2 - 2$ degrees of freedom and noncentrality parameter $\lambda = (\delta_A - \delta_0)/\sigma\sqrt{1/n_1 + 1/n_2}$. Thus, for a specific value of δ_A , β can be determined from the noncentral t-distribution. Tables and charts for determining β from the noncentral t-distribution are

provided by Ferris, Grubbs, and Weaver (1946) and Bowker and Lieberman (1972).

As with the single sample case discussed in Section 4.2.2, application of the noncentral t-distribution is not illustrated. For most situations in practice, however, Equation 4.35 can be used to approximate β for a specified value of δ_A . The problem of specifying a value of $\sigma_1^2 = \sigma_2^2 = \sigma^2$ in Equation 4.35 can be dealt with as in the single sample case. For a given value of σ^2 , the approximation will underestimate β , especially if $n_1 + n_2 - 2 < 30$. Even so, it provides useful information.

Determining sample sizes from the noncentral t-distribution is quite tedious and requires iterative numerical techniques. The tables and charts provided by Ferris, Grubbs, and Weaver (1946) and Bowker and Lieberman (1972) can be used in determining sample sizes. The following approximate procedure proposed by Guenther (1981), however, yields sample sizes which agree quite closely with those computed from exact methods: Use Equation 4.36 or 4.37 to determine an initial value, say n' . Then for the two-sided test, set $n_1 = n_2 = n' + 0.25 (z_{1-\alpha/2})^2$, and for the one-sided test, set $n_1 = n_2 = n' + 0.25 (z_{1-\alpha})^2$.

4.3.2.2 Comparison of Two Means, Assuming Unequal Variances

In some situations, it is not reasonable to assume that $\sigma_1^2 = \sigma_2^2$. When this happens, the test statistic t , given by Equation 4.40, cannot be used to test the hypotheses stated in Equation 4.38, 4.42, or 4.44. Because there is not a common variance, the sample variances are not pooled; however, S_1^2 and S_2^2 provide unbiased estimates of σ_1^2 and σ_2^2 , respectively. Thus, $S_1^2/n_1 + S_2^2/n_2$ gives an unbiased estimate of the variance of $\bar{X}_1 - \bar{X}_2$. Consequently, if σ_1^2 and σ_2^2 are replaced with S_1^2 and S_2^2 in Equation 4.29, the resulting test statistic, t' , is given by

$$t' = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{[(S_1^2/n_1) + (S_2^2/n_2)]^{1/2}} \quad (4.46)$$

However, t' does not follow the Student's t-distribution when H_0 is true; that is, when $\mu_1 - \mu_2 = \delta_0$.

This particular hypothesis testing application is referred to in the statistical literature as the Behrens-Fisher problem. From the 1920s to the present, much research has been done on the distribution of t' . Many of the publications on this subject provide special tables for determining probabilities and percentiles. Some commonly referenced publications with tables include Fisher and Yates (1957), Aspin (1949), and Trickett et al. (1956). The use of these tables is not illustrated. Instead, an approximate procedure due to Satterthwaite (1946) is presented, which may be adequate for most applications.

For the approximate procedure, Equation 3.32 is used to compute the approximate degrees of freedom ν for t' . This formula, repeated here for convenience, is

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}} \quad (4.47)$$

To be conservative, the computed value is always rounded down to an integer value. The Student's t -distribution with ν degrees of freedom is then used to approximate the distribution of t' when $\mu_1 - \mu_2 = \delta_0$.

Thus, to test the two-sided hypothesis in Equation 4.38 at the α level of significance, reject H_0 if either

$$t' \leq -t_{1-\alpha/2}(\nu)$$

or (4.48)

$$t' \geq t_{1-\alpha/2}(\nu)$$

For the one-sided hypothesis in Equation 4.42, reject H_0 if

$$t' \geq t_{1-\alpha}(\nu) \quad (4.49)$$

For the one-sided hypothesis in Equation 4.44, reject H_0 if

$$t' \leq -t_{1-\alpha}(\nu) \quad (4.50)$$

Example 4.13 Suppose that a rapid but imprecise method has been developed for measuring the percent uranium in UO_2 powder. Eight powder samples from a given batch are analyzed by this method. Four samples are analyzed by the standard method, which is precise but very time consuming. It is of interest to test whether the rapid method either overestimates or underestimates the percent uranium.

Assume that μ_1 and σ_1^2 are the mean and variance of the individual analytical results using the rapid method, and μ_2 and σ_2^2 are the mean and

variance for the standard method. The rapid, imprecise method has a larger variance than the standard method. Thus, assume that $\sigma_1^2 \neq \sigma_2^2$ and test the hypothesis

$$H_0: \mu_1 - \mu_2 = 0$$

against the alternative

$$H_A: \mu_1 - \mu_2 = \delta_A \neq 0$$

at the $\alpha = 0.05$ level of significance.

The sample data in percent uranium are as follows

Rapid method X_1	Standard method X_2
87.71	87.78
87.19	87.53
87.64	87.34
86.22	87.55
87.10	
86.89	
86.36	
87.00	

The necessary quantities computed from the data are:

$$n_1 = 8 \quad \bar{X}_1 = 87.0138 \quad S_1^2 = (0.5330)^2$$

$$n_2 = 4 \quad \bar{X}_2 = 87.5500 \quad S_2^2 = (0.1802)^2$$

Substituting these values into Equations 4.46 and 4.47, the test statistic is

$$t' = \frac{(87.0138 - 87.55) - 0}{\{[(0.5330)^2/8] + [(0.1802)^2/4]\}^{1/2}} = -2.57$$

and the approximate degrees of freedom are

$$v = \frac{\left[\frac{(0.5330)^2}{8} + \frac{(0.1802)^2}{4} \right]^2}{\frac{[(0.5330)^2/8]^2}{7} + \frac{[(0.1802)^2/4]^2}{3}} = 9.42 \cong 9$$

From Table A5, $t_{0.975}(9) = 2.26$. Because $-2.57 < -2.26$, reject H_0 and conclude that the quick method underestimates the percent uranium in UO_2 powder relative to the standard method. Methods from Section 3.5.2 can be used to construct a confidence interval for the difference ($\mu_1 - \mu_2$).

When $\mu_1 - \mu_2 = \delta_A \neq \delta_0$, the distribution of t' is unknown, and no tables or charts are available to aid in constructing power curves or OC curves. If $v \geq 30$, however, Equation 4.35 can be used to approximate β . This approximation improves as n_1 and n_2 approach ∞ , but even when n_1 and n_2 are not large, this approximation is quite useful in practice.

To approximate required sample sizes, use of Equation 4.36 or 4.37 with estimates substituted for σ_1^2 and σ_2^2 is recommended. This approximation provides valuable guidance in planning an experiment. The resulting sample size estimate, however, should be increased to help compensate for the inadequacy of the approximation.

4.3.3 Comparison of Two Variances

Assume that independent random samples of size n_1 and n_2 are taken from the normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, where μ_1 , μ_2 , σ_1^2 , and σ_2^2 are unknown. It is of interest to test the hypothesis

$$H_0: \sigma_1^2/\sigma_2^2 = 1$$

against the alternative

(4.51)

$$H_A: \sigma_1^2/\sigma_2^2 = \lambda \neq 1$$

at the α level of significance.

If H_0 is true, then $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and the sample variances S_1^2 and S_2^2 provide independent estimates of the common variance σ^2 . Thus, if H_0 is true, the ratio

$$F = \frac{S_1^2}{S_2^2} \quad (4.52)$$

has an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom (Section 3.5.1). Thus, H_0 is rejected if either

$$F \leq F_{\alpha/2}(n_1 - 1, n_2 - 1)$$

or

(4.53)

$$F \geq F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$$

Because Table A6 gives only the upper percentile points of the F-distribution, the value of $F_{\alpha/2}(n_1 - 1, n_2 - 1)$ is determined from the relationship

$$F_{\alpha/2}(n_1 - 1, n_2 - 1) = \frac{1}{F_{1-\alpha/2}(n_2 - 1, n_1 - 1)} \quad (4.54)$$

To test the one-sided hypothesis

$$H_0: \sigma_1^2/\sigma_2^2 \leq 1$$

against the alternative

(4.55)

$$H_A: \sigma_1^2/\sigma_2^2 = \lambda > 1$$

at the α level of significance, reject H_0 if

$$F \geq F_{1-\alpha}(n_1 - 1, n_2 - 1) \quad (4.56)$$

Similarly, to test the hypothesis

$$H_0: \sigma_1^2/\sigma_2^2 \geq 1$$

against the alternative

(4.57)

$$H_A: \sigma_1^2/\sigma_2^2 = \lambda < 1$$

at the α level of significance, reject H_0 if

$$F \leq F_{\alpha}(n_1 - 1, n_2 - 1) \quad (4.58)$$

Note that any one-sided test can be formulated in terms of the hypotheses and test given by Equations 4.55 and 4.56 by designating which population has variance σ_1^2 and which has variance σ_2^2 . In fact, many statistics texts advocate this convention, and it is commonly used in practice. Thus, the hypotheses and test given by Equations 4.57 and 4.58 are not necessary and are presented here only for completeness.

Bowker and Lieberman (1972) give charts showing the OC curves for the above tests for $\alpha = 0.05$ and $\alpha = 0.10$, assuming that $n_1 = n_2 = n$; however, β can be computed directly for specific values of λ , α , n_1 , and n_2 . For the two-sided test,

$$\beta = \Pr \left[F < \frac{1}{\lambda} F_{1-\alpha/2}(n_1 - 1, n_2 - 1) \right] - \Pr \left[F < \frac{1}{\lambda} F_{\alpha/2}(n_1 - 1, n_2 - 1) \right] \quad (4.59)$$

For the one-sided test given by Equations 4.55 and 4.56,

$$\beta = \Pr \left[F < \frac{1}{\lambda} F_{1-\alpha}(n_1 - 1, n_2 - 1) \right] \quad (4.60)$$

Even with extensive tables of the F-distribution, application of Equations 4.59 and 4.60 requires interpolation. Thus, the values of β are approximate but sufficiently accurate for most applications.

The sample sizes n_1 and n_2 required to achieve a specified value of $1 - \beta$ at a given value of λ can be determined in a "trial-and-error" fashion. This involves calculating β or $1 - \beta$ for some possible values of n_1 and n_2 . This continues until the desired value of β or $1 - \beta$ is found. Then the corresponding values of n_1 and n_2 are the required sample sizes.

Example 4.14 In Example 4.13, the variance of the rapid method is said to be larger than the variance of the standard method. That is, $\sigma_1^2 > \sigma_2^2$. Do the sample results substantiate this at the 0.10 level of significance?

To answer this question, test the hypothesis

$$H_0: \sigma_1^2 / \sigma_2^2 \leq 1$$

against the alternative

$$H_A: \sigma_1^2/\sigma_2^2 = \lambda > 1$$

at the $\alpha = 0.10$ level of significance. The sample results are

$$n_1 = 8 \quad S_1^2 = (0.5330)^2$$

$$n_2 = 4 \quad S_2^2 = (0.1802)^2$$

The test statistic is

$$F = \frac{(0.5330)^2}{(0.1802)^2} = 8.75$$

From Table A6, $F_{0.90}(7, 3) = 5.27$. Because $8.75 > 5.27$, reject H_0 and conclude that σ_1^2 is larger than σ_2^2 .

For the sample sizes $n_1 = 8$ and $n_2 = 4$, determine the power of this test for values of λ in the range $1 \leq \lambda \leq 5$. Equation 4.60 and Table A6 are used. The results are summarized below, and the reader is encouraged to verify these values.

λ	$(1/\lambda)F_{0.90}(7,3)$	β	Power = $1 - \beta$
1.0	5.27	0.90	0.10
2.0	2.64	0.74	0.26
3.0	1.76	0.62	0.38
4.0	1.32	0.53	0.47
5.0	1.05	0.46	0.54

To illustrate the procedure, consider the case $\lambda = 3$. From the previous results, $F_{0.90}(7,3) = 5.27$. Thus,

$$(1/\lambda)F_{0.90}(7,3) = (1/3)(5.27) = 1.76$$

Entering Table A6 with 7 and 3 degrees of freedom, we note that $F_{0.50}(7,3) = 1.15$ and $F_{0.75}(7,3) = 2.43$. Using linear interpolation, the value $F = 1.76$ is approximately the 100th percentile of the F-distribution with 7 and 3 degrees of freedom, where p is approximated as

$$p \cong 0.75 - \frac{(2.43 - 1.76)}{(2.43 - 1.15)} (0.75 - 0.50) = 0.62$$

That is, $1.76 \cong F_{0.62}(7,3)$, and Equation 4.60 becomes

$$\beta = \Pr[F < 1.76] \cong 0.62$$

This test is not very powerful for the values of λ considered. Determine values of n_1 and n_2 such that $1 - \beta = 0.80$ when $\lambda = 5$, subject to the constraint $n_1 = 2n_2$. Begin the trial and error procedure with $n_1 = 10$ and $n_2 = 5$. The procedure for computing β is the same as for the above table. That is, for $n_1 = 10$ and $n_2 = 5$, the value $F_{0.90}(9,4) = 3.94$ is obtained from Table A6, and since $\lambda = 5$,

$$(1/\lambda)F_{0.90}(9,4) = (1/5)(3.94) = 0.79$$

Then entering Table A6 with 9 and 4 degrees of freedom, we note that all values in the body of the table are greater than 1.10. Thus, Equation 4.54 is used to explore the lower tail of the distribution, where

$$F_{\alpha}(9,4) = 1/F_{1-\alpha}(4,9)$$

Entering Table A6 with 4 and 9 degrees of freedom, $F_{0.50}(4,9) = 0.906$ and $F_{0.75}(4,9) = 1.63$. Thus,

$$F_{0.50}(9,4) = 1/0.906 = 1.10$$

and

$$F_{0.25}(9,4) = 1/1.63 = 0.61$$

Then, using linear interpolation, the value $F = 0.79$ is approximately the 100 p^{th} percentile of the F-distribution with 9 and 4 degrees of freedom, where

$$p \cong 0.50 - \frac{(1.10 - 0.79)}{(1.10 - 0.61)} (0.50 - 0.25) = 0.34$$

That is, $0.79 = F_{0.34}(9,4)$, and Equation 4.60 becomes

$$\beta = \Pr[F < 0.79] \cong 0.34$$

The results are summarized below. Sample sizes of $n_1 = 14$ and $n_2 = 7$ satisfy the specified requirements.

n_1, n_2	$F_{0.90}(n_1 - 1, n_2 - 1)$	$(1/5)F_{0.90}(n_1 - 1, n_2 - 1)$	β	$1 - \beta$
10,5	3.94	0.79	0.34	0.66
12,6	3.28	0.66	0.26	0.74
14,7*	2.89	0.58	0.20	0.80
16,8	2.63	0.53	0.14	0.86

*Values calculated by interpolation.

4.4 COMPARISON OF TWO MEANS USING PAIRED OBSERVATIONS

A special case of the comparison of two means occurs when the observations from the two distributions of interest are collected in pairs. The observations in each pair, say (x_{1j}, x_{2j}) , are taken under homogeneous conditions, but these conditions may change from one pair to another. This is the simplest case of the randomized complete block experimental design discussed in Section 7.2.2, where in this case the pairs are the "blocks" described there.

To illustrate this concept, suppose that we are interested in comparing two balances used for weighing UO_2 powder containers. If several containers are randomly selected from those on inventory, and half are randomly assigned to be weighed on balance 1 and half on balance 2, and then one of the independent sample tests of Section 4.3.2 is applied, the results of the test could be misleading. This is because the difference between the two sample means includes a possible difference in the means of the true container weights in the two samples, as well as a possible difference in weights due to the two balances.

A better experimental procedure is to collect the data in pairs. That is, weigh each container on both scales and record the **paired observation** on the j^{th} container as (x_{1j}, x_{2j}) , where x_{1j} and x_{2j} are the weights observed from balances 1 and 2, respectively. The test procedure then consists of analyzing the **differences** $d_j = x_{1j} - x_{2j}$. If there is no difference between the two scales, then the population mean of the differences is zero. Proceeding in this manner, the sample mean of the d_j 's does not include

differences among the true container weights. Thus, the resulting test is more sensitive to the difference between the two balances than the independent sample techniques of Section 4.3.2. This concept is discussed further after the details of the testing procedure and a worked example are presented.

Paired observations are quite common in nuclear material accounting applications. For example, an operator and an inspector measure the same items to verify the operator's reported inventory estimate. Some of the applications where paired observations are used, however, require more sophisticated statistical methods than those presented in this section. Discussion of such cases and the appropriate statistical methods to use are provided in Chapter 17.

Assume that there are two random variables of interest, say X_1 and X_2 , with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 , where a comparison of μ_1 and μ_2 is desired. By taking paired observations, if the pairing is effective, a positive correlation ρ between X_1 and X_2 is introduced. This is discussed in detail after Example 4.15. Thus, the random variable $D = X_1 - X_2$ has mean $\mu_D = \mu_1 - \mu_2$ and variance $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$ in accordance with Equations 2.122 and 2.123. To construct a test which compares μ_1 and μ_2 , assume that the differences $D_j = (X_{1j} - X_{2j})$, $j = 1, 2, \dots, n$, are a random sample from the $N(\mu_D, \sigma_D^2)$ distribution, with μ_D and σ_D^2 unknown. It follows that testing the hypothesis that $(\mu_1 - \mu_2) = \delta_0$ is equivalent to testing the hypothesis

$$H_0: \mu_D = \delta_0$$

against the alternative

$$H_A: \mu_D = \delta_A \neq \delta_0 \quad (4.61)$$

In many applications δ_0 is chosen to be zero, so that H_0 is the hypothesis that $\mu_1 = \mu_2$. The test is conducted by utilizing the single sample techniques of Section 4.2.2, where the sample statistics

$$\bar{D} = \sum_{i=1}^n D_i/n$$

and

$$S_D^2 = \sum_{i=1}^n (D_i - \bar{D})^2 / (n-1)$$

are used to compute the test statistic

$$t = \frac{\bar{D} - \delta_0}{S_D/\sqrt{n}} \quad (4.62)$$

H_0 is rejected at the α level of significance if either

$$t \leq -t_{1-\alpha/2}(n-1)$$

or

$$t \geq t_{1-\alpha/2}(n-1)$$

(4.63)

To test the one-sided hypothesis,

$$H_0: \mu_D \leq \delta_0$$

against the alternative

(4.64)

$$H_A: \mu_D = \delta_A > \delta_0$$

reject H_0 at the α level of significance if

$$t \geq t_{1-\alpha}(n-1) \quad (4.65)$$

Similarly, to test the hypothesis

$$H_0: \mu_D \geq \delta_0$$

against the alternative

(4.66)

$$H_A: \mu_D = \delta_A < \delta_0$$

reject H_0 at the α level of significance if

$$t \leq -t_{1-\alpha}(n-1) \quad (4.67)$$

The discussions in Section 4.2.2 concerning power curves, OC curves, and required sample sizes apply to this procedure and thus are not repeated here. Note, however, that the sample size n presented there corresponds to the number of pairs in this procedure. Thus, in determining the total cost of measurements or experimentation, the total size of the experiment is $2n$.

Example 4.15 Twenty sintered UO_2 pellets were randomly selected from a given batch. Each pellet was split into two parts. One part was randomly selected from each pellet and analyzed for percent uranium by the gravimetric method. The other part was stored for a specified length of time and then analyzed by the same method. Let x_{1j} and x_{2j} be the percent uranium observed for the j^{th} pellet from the initial analysis and the later analysis, respectively. The data below are x_{1j} , x_{2j} , and the difference $d_j = (x_{1j} - x_{2j})$ in percent uranium.

x_1	x_2	d	x_1	x_2	d
88.067	88.063	0.004	88.101	88.063	0.038
88.118	88.072	0.046	88.023	88.011	0.012
88.090	88.045	0.045	88.082	88.070	0.012
88.118	88.125	-0.007	88.038	88.016	0.022
88.085	88.106	-0.021	88.062	88.040	0.022
88.105	88.104	0.001	88.107	88.067	0.040
88.113	88.129	-0.016	88.141	88.116	0.025
88.044	88.041	0.003	88.095	88.074	0.021
88.026	88.021	0.005	88.065	88.045	0.020
88.088	88.063	0.025	88.112	88.134	-0.022

Was there a significant decrease in percent uranium over the specified time period? To answer this question, test the hypothesis

$$H_0: \mu_D \leq 0$$

against the alternative

$$H_A: \mu_D = \delta_A > 0$$

at the $\alpha = 0.05$ level of significance. From the sample data, compute the quantities

$$\bar{D} = 0.01375$$

$$S_D = 0.02056$$

and

$$t = \frac{0.01375 - 0}{0.02056/\sqrt{20}} = 2.99$$

From Table A5, $t_{0.95}(19) = 1.73$. Since $2.99 > 1.73$, reject H_0 and conclude that there was a decrease in the percent uranium content of the pellets over the specified time period.

The formula for the variance of the difference $D = X_1 - X_2$ provides some insight into the circumstances where pairing is effective. From the discussion in Section 2.7.1 and from Equation 2.123, the variance of $D = X_1 - X_2$ is given by

$$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

where

$$\sigma_{12} = E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

is the covariance of X_1 and X_2 . When pairing, the pairs should be chosen such that if X_1 is high, so is X_2 . Then if $X_1 - \mu_1$ is positive, $X_2 - \mu_2$ is usually positive, and their product $(X_1 - \mu_1)(X_2 - \mu_2)$ is usually positive. Similarly, if $(X_1 - \mu_1)$ is negative, $(X_2 - \mu_2)$ is usually negative, and their product is usually positive. Since σ_{12} is essentially the average of the products $(X_1 - \mu_1)(X_2 - \mu_2)$ over the entire population, σ_{12} will be positive when pairing is successful. This makes σ_D^2 smaller than the sum of the individual variances. That is, when $\sigma_{12} > 0$, it follows that

$$\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} < \sigma_1^2 + \sigma_2^2$$

When pairing is successful, σ_D^2 is substantially smaller than $\sigma_1^2 + \sigma_2^2$. Note also that the correlation between X_1 and X_2 is given by $\rho = \sigma_{12}/\sigma_1\sigma_2$. Thus, ρ will be positive and should be significantly greater than zero when pairing is successful.

Pairing is not always effective because X_1 and X_2 may be poorly correlated with the criterion used to define the pairs, and thus X_1 and X_2 are poorly correlated within the resulting pairs. Fortunately, it is possible from the results of a paired experiment to estimate what the variance of $\bar{X}_1 - \bar{X}_2$ would have been if the experiment had been conducted without pairing. With paired samples of size n , the variance of the mean difference $\bar{D} = \bar{X}_1 - \bar{X}_2$ is σ_D^2/n , where σ_D^2 is the variance of the population of paired differences. With two unpaired groups, assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$ the variance of $\bar{X}_1 - \bar{X}_2$ is $2\sigma^2/n$. Thus, the quantities to be compared are σ_D^2 and $2\sigma^2$.

Using the data from a paired experiment, the sample variance S_D^2 is an unbiased estimator of σ_D^2 . Estimation of σ^2 from the paired data, however, is slightly more difficult. One possibility is to analyze the paired data by the method given in Section 4.3.2.1 for two independent samples using the pooled variance S_p^2 , given by Equation 4.39, as an estimate of σ^2 . For a comparison which utilizes the $2n$ observed values from a paired experiment, however, this procedure is not quite correct. This is because if the experiment had been conducted by randomly assigning half of the $2n$ objects to one "treatment group" and half to the other (in Example 4.15, this would be equivalent to randomly selecting n of the $2n$ pellet halves for immediate analysis and storing the remaining n for later analysis), it is very likely that some pairs would have one member in each treatment group. Thus, a partial pairing occurs without deliberate planning so that X_1 and X_2 are not totally independent. An unbiased estimator of $2\sigma^2$ is not $2S_p^2$ but is given by

$$2\hat{\sigma}^2 = 2S_p^2 - (2S_p^2 - S_D^2)/(2n - 1)$$

From the data of Example 4.15, we compute $S_D^2 = (0.02056)^2 = 0.0004227$, $S_p^2 = 0.0012550$, and

$$2\hat{\sigma}^2 = 2(0.001255) - [2(0.001255) - 0.0004227]/39 = 0.0024565$$

In this example, the pairing has resulted in a much smaller variance of $\bar{X}_1 - \bar{X}_2$; that is, $0.0004227/n$ with pairing compared to $0.0024565/n$ without pairing. In practical terms, this implies that with independent samples, the sample size would have to be increased from 20 pairs to independent samples of size $n = 20(0.0024565)/0.0004227 \cong 116$ each to

give the same variance of $\bar{X}_1 - \bar{X}_2$ as the paired experiment. In this case, the savings in terms of cost, time, and so on, due to pairing are substantial. That is, $232/40 = 5.8$ times as many pellet halves are required without pairing to achieve the same power that results from pairing.

The above comparison overlooks one important point. In the paired experiment, S_D^2 has $n-1$ degrees of freedom, whereas S_p^2 has $2(n-1)$ degrees of freedom. Thus, the t value obtained from Table A5 to be used in the hypothesis test is smaller for the independent samples test than for the paired test. Before comparing S_D^2 with $2\hat{\sigma}^2$, Fisher (1960) advocates multiplying each estimate by the factor $(\nu + 3)/(\nu + 1)$, where ν is $n-1$ for S_D^2 and $2(n-1)$ for the $\hat{\sigma}^2$. In the above example, this gives $23(0.0004227)/21 = 0.0004630$ and $41(0.0024565)/39 = 0.0025825$. Then making the same comparison as above, the sample size must be increased from 20 pairs to independent samples of size $n = 20(0.0025825)/0.0004630 \cong 112$ each to give the same variance of $\bar{X}_1 - \bar{X}_2$ as the paired experiment. That is, in this example, $224/40 = 5.6$ times as many pellet halves are required without pairing to achieve the same power that results from pairing.

If the criterion used for pairing has no correlation with X_1 and X_2 , a small loss in power of the test, relative to the independent samples test, results from pairing due to the adjustment for degrees of freedom. A substantial loss in power can occur if the criterion for pairing is so poorly chosen that it results in a negative correlation between members of a pair.

Sometimes, in practice, an experimenter with paired data forgets or ignores the fact that they are paired and carries out the hypothesis test as if the two samples were independent. This is a serious mistake if the pairing was effective. In Example 4.15, the value

$$\sqrt{2S_p^2/20} = 0.0112$$

would be used in the denominator of the test statistic rather than

$$\sqrt{S_D^2/20} = 0.0046$$

This mistake negates the advantage of pairing, and a difference between \bar{X}_1 and \bar{X}_2 that is actually statistically significant may be deemed nonsignificant.

4.5 TESTS ON PROPORTIONS AND COUNT RATES

There are situations in practice where the random variable of interest follows either the binomial distribution or the Poisson distribution. These

distributions are introduced in Sections 2.4.2 and 2.4.3, respectively. This section presents exact methods for testing hypotheses about the parameter p of the binomial distribution and the parameter μ of the Poisson distribution. Also, approximate methods are presented for testing hypotheses about these parameters and for comparing the parameters of two binomial distributions using data from independent samples.

4.5.1 Exact Tests on the Binomial Parameter p

In Section 2.4.2, a binomial random variable X is defined as the number of successes observed in n independent trials of an experiment with two possible outcomes, where the probability of a success at each trial is p .

Consider testing the hypothesis

$$H_0: p = p_0$$

against the alternative

(4.68)

$$H_A: p = p_A \neq p_0$$

at the α level of significance. Assume that a random sample of size n is taken from a binomial distribution and that x successes are observed. If H_0 is true, the random variable X has a binomial distribution with parameters n and p_0 . Thus, the rejection region is determined by using Equations 2.20 and 2.3 to find limits L and U such that

$$\begin{aligned} \Pr(L \leq X \leq U | p = p_0) \\ = \sum_{x=L}^U \binom{n}{x} p_0^x (1-p_0)^{n-x} \cong 1 - \alpha \end{aligned} \quad (4.69)$$

H_0 is rejected if either

$$x < L$$

or

(4.70)

$$x > U$$

Because X is a discrete random variable, it is rare in practice that H_0 can be tested at exactly the α level of significance. That is, Equation 4.69

is rarely an equality; however, L and U should be chosen so that they satisfy as closely as possible the criteria

$$\Pr(X < L | p = p_0) = \sum_{x=0}^{L-1} \binom{n}{x} p_0^x (1-p_0)^{n-x} \cong \alpha/2$$

and

$$\begin{aligned} \Pr(X > U | p = p_0) &= \sum_{x=U+1}^n \binom{n}{x} p_0^x (1-p_0)^{n-x} \\ &= 1 - \sum_{x=0}^U \binom{n}{x} p_0^x (1-p_0)^{n-x} \cong \alpha/2 \end{aligned} \quad (4.71)$$

The actual significance level of the test is given by

$$\alpha^* = \Pr(X < L | p = p_0) + \Pr(X > U | p = p_0) \quad (4.72)$$

To evaluate Equations 4.69, 4.71, and others discussed in this section, a computer or calculator that has been programmed to compute binomial probabilities and cumulative binomial probabilities is desirable.

To test the one-sided hypothesis

$$H_0: p \leq p_0$$

against the alternative (4.73)

$$H_A: p = p_A > p_0$$

at approximately the α level of significance, find the limit U such that

$$\alpha^* = \Pr(X > U | p = p_0) \cong \alpha \quad (4.74)$$

H_0 is rejected at the α level of significance if

$$x > U \quad (4.75)$$

To test the hypothesis

$$H_0 : p \geq p_0$$

against the alternative (4.76)

$$H_A : p = p_A < p_0$$

at approximately the α level of significance, find the limit L such that

$$\alpha^* = \Pr(X < L | p = p_0) \cong \alpha \quad (4.77)$$

H_0 is rejected at the α^* level of significance if

$$x < L \quad (4.78)$$

To construct a power curve or an OC curve for the two-sided test, compute β at specified values of p_A , where

$$\beta = \Pr(L \leq X \leq U | p = p_A) \quad (4.79)$$

Note that this is Equation 4.69 with p_A in place of p_0 . For the one-sided hypothesis in Equation 4.73, β is given by

$$\beta = \Pr(X \leq U | p = p_A) \quad (4.80)$$

Similarly, for the one-sided hypothesis in Equation 4.76, β is given by

$$\beta = \Pr(X \geq L | p = p_A) \quad (4.81)$$

The value of β in Equation 4.79, 4.80, or 4.81 is computed from the binomial distribution with parameters n and p_A . The power of the test when $p = p_A$ is $1 - \beta$ in each case.

With the aid of a computer or programmed calculator, the following trial-and-error procedure is adequate for determining the sample size

required to ensure that the power of the test is at a specified level when p takes on a specified value p_A . For specified values of α , β , p_0 , and p_A , this process begins by arbitrarily selecting an initial value of n . The limits L and/or U are determined such that Equation 4.71, 4.74, or 4.77 is satisfied as closely as possible. Then β is computed from Equation 4.79, 4.80, or 4.81. If the computed β is much smaller than the specified value of β , the sample size may be unnecessarily large, and the above steps are repeated with a smaller value of n . If, however, the computed β is larger than the specified value, then the sample size is too small, and the above steps are repeated with a larger value of n . This process continues until a value of n is found where α^* and the computed β are acceptably close to the specified values of α and β .

For the one-sided hypothesis given by Equation 4.73, Jaech (1980) presents a method to determine the required sample size and the rejection limit U for specified values of p_0 , p_A , α , and β . His method is not a trial-and-error procedure, and the results are obtained by using tables of the normal and chi-square distributions together with special tables provided by Stephens (1978). The details of Jaech's method are not presented here.

Example 4.16 Suppose that in Example 2.8 the plant manager recommends that a random sample of 100 containers be taken periodically from the process and weighed. The weights are recorded to the nearest 10 g. The process is said to be in control if no more than 5% of the containers filled by this process have weights that differ from the prespecified filled weight. If p denotes the percent of containers with weight discrepancies, then the hypothesis to be tested is

$$H_0: p \leq 0.05$$

against the alternative

$$H_A: p = p_A > 0.05$$

Assume that a significance level of $\alpha = 0.10$ is specified.

In the most recent sample of 100 containers, seven have weights that differ from the prespecified value. To test the above hypothesis, first consider the pdf of the number of discrepancies, X , when $p = 0.05$ and $n = 100$. Using Equation 2.20, this is

x	$\Pr(X = x p = 0.05)$	x	$\Pr(X = x p = 0.05)$
0	0.0059	9	0.0349
1	0.0312	10	0.0167
2	0.0812	11	0.0072
3	0.1396	12	0.0028
4	0.1781	13	0.0010
5	0.1800	14	0.0003
6	0.1500	15	0.0001
7	0.1060	≥ 16	0.0000
8	0.0649		

To determine the rejection region, note that $\Pr(X > 7) = 0.1280$ and $\Pr(X > 8) = 0.0631$. To be conservative, choose $U = 8$ with $\alpha^* = 0.0631$. That is, if $x > 8$, H_0 is rejected at the 0.0631 level of significance. Since $x = 7$, however, H_0 is not rejected, and we conclude that the process is in control.

Points for constructing either a power curve or an OC curve for this test are obtained by using Equation 4.80 and are tabulated below for some selected values of p_A .

p_A	$\beta = \Pr(X \leq U p = p_A)$	Power = $1 - \beta$
0.010	1.0000	0.0000
0.025	0.9991	0.0009
0.050	0.9369	0.0631
0.075	0.6648	0.3352
0.100	0.3209	0.6791
0.125	0.1088	0.8912
0.150	0.0275	0.9725

Suppose the plant manager decides that the power of the test must be 0.80 (β must be 0.20) when $p = 0.09$. What sample size is required?

The trial-and-error results are as follows:

n	U	α^*	$\beta = \Pr(X \leq U p = 0.09)$	Power = $1 - \beta$
100	8	0.0631	0.4494	0.5006
120	9	0.0786	0.3530	0.6470
140	10	0.0931	0.2761	0.7239
160	11	0.1064	0.2154	0.7846
165	12	0.0714	0.2687	0.7313
170	12	0.0855	0.2314	0.7686
175	12	0.1012	0.1978	0.8022

A plan that reasonably satisfies the plant manager's specifications is to take a sample of size 175 and reject H_0 at the $\alpha^* = 0.1012$ level of significance if $x > 12$. This plan has power 0.8022 when $p = 0.09$.

4.5.2 Approximate Tests on the Binomial Parameter p

This section presents a procedure based upon the normal approximation to the binomial for testing the hypotheses given by Equations 4.68, 4.73, and 4.76. This approximate procedure is adequate when the conditions for applying the normal approximation to the binomial are met for specified values of n and p_0 (Section 2.5.2.2). If these conditions are met and H_0 is true, then the random variable X , the number of successes, is approximately distributed as $N[np_0, np_0(1 - p_0)]$.

To test the hypothesis in Equation 4.68, assume that a random sample of size n is taken from a binomial distribution and that X successes are observed. Compute the test statistic Z , given in terms of either X or \hat{p} as

$$Z = \frac{X - np_0}{[np_0(1 - p_0)]^{1/2}} = \frac{\hat{p} - p_0}{[p_0(1 - p_0)/n]^{1/2}} \quad (4.82)$$

where $\hat{p} = X/n$. Reject H_0 at the α level of significance if either

$$Z \leq -z_{1-\alpha/2}$$

or

$$Z \geq z_{1-\alpha/2}$$

where $z_{1-\alpha/2}$ is obtained from Table A3.

To test the one-sided hypothesis given by Equation 4.73 at the α level of significance, reject H_0 if

$$Z \geq z_{1-\alpha}$$

Similarly, to test the one-sided hypothesis given by Equation 4.76 at the α level of significance, reject H_0 if

$$Z \leq -z_{1-\alpha}$$

Approximations of power and required sample size are derived by noting that when H_0 is false, $p = p_A$ and X has an approximate $N[np_A, np_A(1-p_A)]$ distribution. That is, both the mean and variance of X are different when $p = p_A$ than when $p = p_0$.

Points for constructing power curves or OC curves are obtained for the two-sided test by using the approximation

$$\beta \cong \Phi \left\{ z_{1-\alpha/2} \left[\frac{p_0(1-p_0)}{p_A(1-p_A)} \right]^{1/2} - \frac{p_A - p_0}{\sqrt{p_A(1-p_A)/n}} \right\} - \Phi \left\{ -z_{1-\alpha/2} \left[\frac{p_0(1-p_0)}{p_A(1-p_A)} \right]^{1/2} - \frac{p_A - p_0}{\sqrt{p_A(1-p_A)/n}} \right\} \quad (4.83)$$

For the one-sided hypothesis in Equation 4.73, an approximate value of β is

$$\beta \cong \Phi \left\{ z_{1-\alpha} \left[\frac{p_0(1-p_0)}{p_A(1-p_A)} \right]^{1/2} - \frac{p_A - p_0}{\sqrt{p_A(1-p_A)/n}} \right\} \quad (4.84)$$

For the one-sided hypothesis in Equation 4.76, the approximation is

$$\beta \cong 1 - \Phi \left\{ -z_{1-\alpha} \left[\frac{p_0(1-p_0)}{p_A(1-p_A)} \right]^{1/2} - \frac{p_A - p_0}{\sqrt{p_A(1-p_A)/n}} \right\} \quad (4.85)$$

For the two-sided hypothesis, when approximating the sample size n required to ensure a specified power of the test when $|p_A - p_0| = \Delta$, there are two separate values of p_A to consider: $p_A = p_0 + \Delta$ and $p_A = p_0 - \Delta$. They give different values of $p_A(1-p_A)$ unless $p_0 = 0.50$. In computing sample size, use the value of p_A corresponding to the maximum value of $p_A(1-p_A)$. Then the approximate required sample size is given by

$$n \cong \frac{\left[z_{1-\alpha/2} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_A(1-p_A)} \right]^2}{(p_A - p_0)^2} \quad (4.86)$$

For the one-sided hypothesis in Equation 4.73, use $p_A = p_0 + \Delta$ to compute

$$n \cong \frac{\left[z_{1-\alpha} \sqrt{p_0(1-p_0)} + z_{1-\beta} \sqrt{p_A(1-p_A)} \right]^2}{(p_A - p_0)^2} \quad (4.87)$$

For the one-sided hypothesis in Equation 4.76, use $p_A = p_0 - \Delta$ in Equation 4.87 to compute the required sample size.

Example 4.17 In Example 4.16, note that $n = 100$, $p_0 = 0.05$, and $np_0 = 5$. Thus, the rules of thumb for applying the normal approximation to the binomial (as stated in Section 2.5.2.2) are satisfied. Use the approximate procedure to test the hypothesis

$$H_0: p \leq 0.05$$

against the alternative

$$H_A: p > 0.05$$

at the $\alpha = 0.10$ level of significance.

Recall that there are $x = 7$ containers with weight discrepancies. Thus, $\hat{p} = 7/100 = 0.07$, and the test statistic is computed from Equation 4.82 as

$$Z = \frac{0.07 - 0.05}{[0.05(0.95)/100]^{1/2}} = 0.92$$

From Table A3, $z_{0.90} = 1.28$. Since $0.92 < 1.28$, H_0 is not rejected. We conclude that the process is in control. This agrees with the conclusion reached in Example 4.16.

In Example 4.16, the plant manager specified that the power of the test must be 0.80 when $p = 0.09$. Determine the required sample size for the approximate test.

Substituting $z_{0.90} = 1.28$, $z_{0.80} = 0.84$, $p_0 = 0.05$, and $p_A = 0.09$ into Equation 4.87 gives

$$n \cong \frac{[1.28\sqrt{0.05(0.95)} + 0.84\sqrt{0.09(0.91)}]^2}{(0.09 - 0.05)^2} = 168.6 \cong 169$$

That is, the sample size must be increased from 100 to 169 to satisfy the requirements. This is in reasonably close agreement with the results of Example 4.16, where a sample of $n = 175$ was recommended for the exact test.

4.5.3 Comparing Two Proportions from Independent Samples

Consider the case where there are two binomial populations of interest with probabilities of success p_1 and p_2 , and we wish to test the hypothesis

$$H_0: p_1 = p_2$$

against the alternative

(4.88)

$$H_A: p_1 \neq p_2$$

at the α level of significance.

An exact procedure for testing this hypothesis is given by Nelson (1982). This section presents an approximate procedure based upon the normal approximation to the binomial.

Assume that independent random samples of size n_1 and n_2 have been taken from two binomial distributions, and that $\hat{p}_1 = X_1/n_1$ and $\hat{p}_2 = X_2/n_2$ are the estimators for p_1 and p_2 , where X_1 and X_2 are the number of successes in the two samples. Assume further that the conditions for applying the normal approximation to the binomial are satisfied for n_1 and p_1 and for n_2 and p_2 .

Under these assumptions, if H_0 is true and $p_1 = p_2 = p$, then the random variable

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{[p(1-p)(1/n_1 + 1/n_2)]^{1/2}}$$

is approximately distributed as $N(0, 1)$. An estimator for the common parameter p is given by

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \quad (4.89)$$

The test statistic is then given by

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{[\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)]^{1/2}} \quad (4.90)$$

To test the two-sided hypothesis given by Equation 4.88, H_0 is rejected at the α level of significance if either

$$Z \leq -z_{1-\alpha/2}$$

or

(4.91)

$$Z \geq z_{1-\alpha/2}$$

To test the one-sided hypothesis

$$H_0 : p_1 \leq p_2$$

against the alternative

(4.92)

$$H_A : p_1 > p_2$$

at the α level of significance, reject H_0 if

$$Z \geq z_{1-\alpha} \quad (4.93)$$

Similarly, to test the hypothesis

$$H_0 : p_1 \geq p_2$$

against the alternative

(4.94)

$$H_A : p_1 < p_2$$

at the α level of significance, reject H_0 if

$$Z \leq -z_{1-\alpha} \quad (4.95)$$

Only approximate techniques are available for determining power curves, OC curves, and required sample sizes. These involve using the formula

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \quad (4.96)$$

for the variance of $\hat{p}_1 - \hat{p}_2$.

Approximate power curves and OC curves are constructed by substituting $(p_1 - p_2)$ for $(\delta_A - \delta_0)$, $p_1(1-p_1)$ for σ_1^2 , and $p_2(1-p_2)$ for σ_2^2 in Equation 4.35. Required sample sizes are approximated as follows. For the two-sided test, to control the power of the test at a specified level $1-\beta$ for given values of α and $(p_1 - p_2)$, the approximate required sample sizes are given by

$$n_1 = n_2 \cong \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2} \quad (4.97)$$

For the one-sided tests, the approximate required sample sizes are given by

$$n_1 = n_2 \cong \frac{(z_{1-\alpha} + z_{1-\beta})^2 [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2} \quad (4.98)$$

One difficulty in applying the above approximations is that values of p_1 and p_2 must be specified. The discussions given previously on specifying estimates of unknown parameters for constructing OC curves and determining required sample sizes apply here. If the experimenter has no knowledge at all about possible values of p_1 and p_2 , substituting the value $p_1 = p_2 = 0.5$ in the numerator of Equation 4.97 or 4.98 yields the maximum sample size over all possible values of p_1 and p_2 and thus provides a conservative estimate of the required sample size.

Example 4.18 Suppose that in Example 4.16 random samples of size $n_1 = 100$ and $n_2 = 100$ were taken during two successive periods and that

the numbers of weight discrepancies observed are $x_1 = 4$ and $x_2 = 9$. Test the hypothesis that the unknown true percent of containers having weight discrepancies is the same for both periods. Use the $\alpha = 0.10$ level of significance. That is, test the hypothesis

$$H_0: p_1 = p_2$$

against the alternative

$$H_A: p_1 \neq p_2$$

From the sample information, compute the quantities

$$\hat{p}_1 = 4/100 = 0.04$$

$$\hat{p}_2 = 9/100 = 0.09$$

$$\hat{p} = (4 + 9)/(100 + 100) = 0.065$$

and the test statistic

$$Z = \frac{0.04 - 0.09}{\{0.065(0.935)[(1/100) + (1/100)]\}^{1/2}} = -1.43$$

From Table A3, $z_{0.95} = 1.65$. Since $-1.65 < -1.43 < 1.65$, do not reject H_0 at the 0.10 level. There is insufficient evidence to conclude that the proportion of containers with weight discrepancies is different for the two periods.

How large must the sample sizes be to ensure that the power of the test is 0.90 when $|p_1 - p_2| \geq 0.05$? To answer this, use Equation 4.97 and assume that the process is in control, and historical data indicate that $p_1(1-p_1) \cong p_2(1-p_2) = 0.05(0.95) = 0.0475$. This gives approximate required sample sizes

$$n_1 = n_2 \cong \frac{(1.65 + 1.28)^2(0.0475 + 0.0475)}{(0.05)^2} = 326.23 \cong 327$$

That is, samples of size $n_1 = n_2 = 327$ are required to meet the specifications.

A few calculations of this type illustrate the fact that large samples are required to detect small differences between two proportions.

4.5.4 Exact Tests on the Poisson Parameter

The Poisson distribution is introduced in Section 2.4.3 with pdf given by Equation 2.25. The random variable X is the frequency of occurrence of some event within some observed time, area, volume, and so on. For example, a nondestructive assay might consist of counting the number of gamma emissions from a sample of material within a specified period of time. If the count rate is λ emissions per minute, the number of particles counted in a 5-min interval follows the Poisson distribution with parameter $\mu = 5\lambda$. In general, let λ be the rate of occurrence per unit time (area, volume, etc.) of some event, and let T denote the number of units of time (area, volume, etc.) over which the number of events is counted. Then the observed count is assumed to follow a Poisson distribution with parameter $\mu = \lambda T$.

Consider testing the hypothesis that the count rate λ is equal to a specified value λ_0 , when the random variable X is the number of events counted over T units of time (area, volume, etc.). Thus, the hypothesis to be tested is

$$H_0: \lambda = \lambda_0$$

against the alternative (4.99)

$$H_A: \lambda = \lambda_A \neq \lambda_0$$

Let α denote the desired level of significance.

If H_0 is true, X has a Poisson distribution with parameter $\mu = \lambda_0 T$. Thus, the rejection region is determined by using Equations 2.25 and 2.3 to find limits L and U such that

$$\Pr(L \leq X \leq U | \mu = \lambda_0 T) = \sum_{x=L}^U \frac{(\lambda_0 T)^x}{x!} e^{-\lambda_0 T} \cong 1 - \alpha \quad (4.100)$$

Then H_0 is rejected if either

$$x < L$$

or (4.101)

$$x > U$$

where x is the observed count.

Because X is a discrete random variable, it is rare in practice that H_0 can be tested at exactly the α level of significance; however, L and U should be chosen such that they satisfy as closely as possible the criteria

$$\Pr(X < L | \mu = \lambda_0 T) = \sum_{x=0}^{L-1} \frac{(\lambda_0 T)^x}{x!} e^{-\lambda_0 T} \cong \alpha/2$$

and

(4.102)

$$\begin{aligned} \Pr(X > U | \mu = \lambda_0 T) &= \sum_{x=U+1}^{\infty} \frac{(\lambda_0 T)^x}{x!} e^{-\lambda_0 T} \\ &= 1 - \sum_{x=0}^U \frac{(\lambda_0 T)^x}{x!} e^{-\lambda_0 T} \cong \alpha/2 \end{aligned}$$

The actual significance level of the test is given by

$$\alpha^* = \Pr(X < L | \mu = \lambda_0 T) + \Pr(X > U | \mu = \lambda_0 T) \quad (4.103)$$

In applying Equations 4.100 and 4.102, a computer or calculator that is programmed to compute Poisson probabilities and cumulative Poisson probabilities is desirable.

To test the one-sided hypothesis

$$H_0: \lambda \leq \lambda_0$$

against the alternative

(4.104)

$$H_A: \lambda = \lambda_A > \lambda_0$$

at approximately the α level of significance, find the limit U such that

$$\alpha^* = \Pr(X > U | \mu = \lambda_0 T) \cong \alpha \quad (4.105)$$

H_0 is rejected at the α^* level of significance if

$$x > U \quad (4.106)$$

To test the hypothesis

$$H_0: \lambda \geq \lambda_0$$

against the alternative

$$(4.107)$$

$$H_A: \lambda = \lambda_A < \lambda_0$$

at approximately the α level of significance, find the limit L such that

$$\alpha^* = \Pr(X < L | \mu = \lambda_0 T) \cong \alpha \quad (4.108)$$

H_0 is rejected at the α^* level of significance if

$$x < L \quad (4.109)$$

Either a power curve or an OC curve is constructed for the two-sided test by computing β at specified values of λ_A , where

$$\beta = \Pr(L \leq X \leq U | \mu = \lambda_A T) \quad (4.110)$$

Note that this is Equation 4.100 with λ_A in place of λ_0 . For the one-sided hypothesis in Equation 4.104, β is given by

$$\beta = \Pr(x \leq U | \mu = \lambda_A T) \quad (4.111)$$

Similarly, for the one-sided hypothesis in Equation 4.107, β is given by

$$\beta = \Pr(X \geq L | \mu = \lambda_A T) \quad (4.112)$$

The value of β in Equation 4.110, 4.111, or 4.112 is computed from the Poisson distribution with parameter $\mu = \lambda_A T$, and the power in each case is $1 - \beta$.

Determining the value of T (the number of units of time, area, volume, and so on) required to ensure that the power of the test is at a specified level when λ takes on a specified value λ_A is a trial-and-error process. For specified values of α , β , λ_0 , and λ_A , this process begins by arbitrarily selecting an initial value of T . The limits L and/or U are determined such

that Equation 4.102, 4.105, or 4.108 is satisfied as closely as possible. Then β is computed from Equation 4.110, 4.111, or 4.112. If the computed β is much smaller than the specified value of β , then T may be unnecessarily large, and the above steps are repeated with a smaller value of T . If, however, the computed β is larger than the specified value of β , then T is too small, and the above steps are repeated with a larger value of T . This process continues until a value of T is found where α^* and the computed β are acceptably close to the specified values of α and β .

Example 4.19 A fuel fabrication plant is required to comply with EPA's radon standard of 5 pCi/liter when discharging its liquid wastes. To monitor the liquid discharge, a 1-liter sample is taken periodically and prepared such that daughter products in secular equilibrium with radon are collected on a small filter. If the discharge is precisely at the EPA standard, the filter will emit gamma particles at the rate of 11 per min. A laboratory counting instrument is used to count gamma emissions from the filter. The counting instrument is only 50% effective, however, so that the expected count rate is 5.5 emissions per min when the liquid discharge is at the EPA standard.

To demonstrate compliance, an instrument count rate less than or equal to 3 emissions per min is desirable. To monitor this, a test of the hypothesis

$$H_0: \lambda \leq 3.0$$

against the alternative

$$H_A: \lambda > 3.0$$

is to be performed at the $\alpha = 0.10$ level of significance each time a sample is analyzed. If, however, the EPA standard of 5.5 emissions per min is reached or exceeded, this must be detected with a probability 0.90 (i.e., $\beta = 0.10$).

Determine T , the number of minutes the filter must be counted, and describe the hypothesis test that satisfies the specified requirements.

The trial-and-error procedure for determining the required value of T begins with $T = 1$. Equation 4.105 is used to determine U and α^* , and Equation 4.111 is used to compute β . The results are summarized below.

T	$\lambda_0 T$	$\lambda_A T$	U	$\alpha^* =$ $\Pr(X > U \mu = \lambda_0 T)$	$\beta =$ $\Pr(X \leq U \mu = \lambda_A T)$
1.0	3.0	5.50	5	0.084	0.615
2.0	6.0	11.00	9	0.084	0.587
3.0	9.0	16.50	13	0.074	0.236
4.0	12.0	22.00	16	0.101	0.117
4.5	13.5	24.75	18	0.092	0.100
5.0	15.0	27.50	20	0.083	0.086

A procedure that satisfies the specified requirements is to “count” the filter for 4.5 min and reject H_0 if $x > 18$ emissions are counted in the 4.5-min interval.

4.5.5 Approximate Tests on the Poisson Parameter

When $\mu \geq 50$, the Poisson distribution is satisfactorily approximated by the $N(\mu, \mu)$ distribution (recall from Section 2.4.3 that the mean and variance of a Poisson distribution are equal). Thus, under H_0 in Equation 4.99, the test statistic

$$Z = \frac{X - \lambda_0 T}{\sqrt{\lambda_0 T}} \quad (4.113)$$

is approximately distributed as $N(0,1)$.

To test the two-sided hypothesis in Equation 4.99, H_0 is rejected at the α level of significance if

$$Z \leq -z_{1-\alpha/2}$$

or (4.114)

$$Z \geq z_{1-\alpha/2}$$

where $z_{1-\alpha/2}$ is obtained from Table A3.

The null hypothesis H_0 in Equation 4.104 is rejected at the α level of significance if

$$Z \geq z_{1-\alpha} \quad (4.115)$$

Similarly, H_0 in Equation 4.107 is rejected at the α level of significance if

$$Z \leq -z_{1-\alpha} \quad (4.116)$$

Approximate power curves or OC curves are constructed for the above tests by computing β from the $N(0,1)$ distribution for specified values of λ_A . For the two-sided test, β is approximated by

$$\beta \cong \phi \left[\frac{\sqrt{T}(\lambda_0 - \lambda_A) + z_{1-\alpha/2} \sqrt{\lambda_0}}{\sqrt{\lambda_A}} \right] - \phi \left[\frac{\sqrt{T}(\lambda_0 - \lambda_A) - z_{1-\alpha/2} \sqrt{\lambda_0}}{\sqrt{\lambda_A}} \right] \quad (4.117)$$

For the one-sided hypothesis given by Equation 4.104, β is approximated by

$$\beta \cong \phi \left[\frac{\sqrt{T}(\lambda_0 - \lambda_A) + z_{1-\alpha} \sqrt{\lambda_0}}{\sqrt{\lambda_A}} \right] \quad (4.118)$$

For the one-sided hypothesis given by Equation 4.107, β is approximated by

$$\beta \cong 1 - \phi \left[\frac{\sqrt{T}(\lambda_0 - \lambda_A) - z_{1-\alpha} \sqrt{\lambda_0}}{\sqrt{\lambda_A}} \right] \quad (4.119)$$

In each case, the power of the test at λ_A is $1 - \beta$.

In approximating the value of T required for the two-sided test to ensure that β is at a specified level when $|\lambda_A - \lambda_0| = \Delta$, there are two separate values of λ_A to consider: $\lambda_A = \lambda_0 + \Delta$ and $\lambda_A = \lambda_0 - \Delta$. Because the variance of X is $\lambda_A T$, the value $\lambda_A = \lambda_0 + \Delta$ yields the maximum value of the variance and is used to determine the required value of T for the two-sided test from the formula

$$T \cong \frac{(z_{1-\alpha/2} \sqrt{\lambda_0} + z_{1-\beta} \sqrt{\lambda_A})^2}{(\lambda_A - \lambda_0)^2} \quad (4.120)$$

For either of the one-sided hypotheses in Equations 4.104 and 4.107, the approximate required value of T is given by

$$T \cong \frac{(z_{1-\alpha} \sqrt{\lambda_0} + z_{1-\beta} \sqrt{\lambda_A})^2}{(\lambda_A - \lambda_0)^2} \quad (4.121)$$

Note that in Equation 4.104, $\lambda_A > \lambda_0$ so that $\lambda_A = \lambda_0 + \Delta$ is used in Equation 4.121; whereas in Equation 4.107, $\lambda_A < \lambda_0$ so that $\lambda_A = \lambda_0 - \Delta$ is used in Equation 4.121.

Example 4.20 A gamma counting instrument used for nondestructive assays is tested periodically for accuracy. A standard that is known to emit gamma particles at the rate of 20,000 per min is counted by the instrument for 5 min. The efficiency of the instrument is 50%, so that if the instrument is in proper adjustment, its expected effective count rate is 10,000 emissions per min.

In a 5-min test, the observed count was 49,532. Test the hypothesis

$$H_0: \lambda = 10,000 \text{ per min}$$

against the alternative

$$H_A: \lambda = \lambda_A \neq 10,000 \text{ per min}$$

at the $\alpha = 0.10$ level of significance, and calculate the points for an OC curve for this test over the range $9,800 \leq \lambda_A \leq 10,200$.

Equation 4.113 is used to compute the test statistic

$$Z = \frac{49,532 - (10,000)(5)}{\sqrt{(10,000)(5)}} = -2.09$$

From Table A3, $z_{0.95} = 1.65$. Since $-2.09 < -1.65$, reject H_0 and conclude that the count rate is lower than 10,000 emissions per min. Thus, the instrument is in need of adjustment.

To compute the points for either a power curve or an OC curve for the above test over the range $9,800 \leq \lambda_A \leq 10,200$, Equation 4.117 is used,

and the results are summarized in the table below. The reader is encouraged to verify several of these values of β .

λ_A	β	Power	λ_A	β	Power
9,800	0.002	0.998	10,050	0.696	0.304
9,850	0.043	0.957	10,100	0.281	0.719
9,900	0.278	0.722	10,150	0.046	0.954
9,950	0.571	0.429	10,200	0.003	0.997
10,000	0.900	0.100			

Suppose the plant manager requires that $\beta = 0.10$ when $|\lambda_A - 10,000| = 100$. How many minutes must the instrument test be run to meet this requirement? Using Equation 4.120 with $\lambda_A = 10,100$ gives

$$T \cong \frac{(1.65\sqrt{10,000} + 1.28\sqrt{10,100})^2}{(10,000 - 10,100)^2} = 8.62$$

Thus, the length of the test must be increased from 5 min to 8 min and 38 sec to meet the requirements.

4.6 SEQUENTIAL TESTS

All of the tests considered thus far in this chapter are based upon samples of a predetermined, fixed size. In this section, a procedure is given for testing the hypothesis

H_0 : the distribution being sampled has pdf $f_0(x)$

against the alternative

(4.122)

H_A : the distribution being sampled has pdf $f_A(x)$

where the sample size is not fixed in advance. The procedure is called a **sequential probability ratio test**, or **SPRT**, and is the topic of texts by Wald (1947) and Wetherill (1975). The reader is referred to these texts for more complete and detailed discussions of the subject.

A sequential test proceeds by taking a sequence of independent observations on a random variable X , denoted by x_1, x_2, \dots , one at a time, and making one of three decisions after each observation. For the m^{th} observation, $m = 1, 2, \dots$, the decisions are as follows:

1. Continue sampling by taking an $(m + 1)^{\text{st}}$ observation if

$$B < \frac{f_A(x_1) f_A(x_2) \dots f_A(x_m)}{f_0(x_1) f_0(x_2) \dots f_0(x_m)} < A \quad (4.123)$$

2. Stop sampling and accept H_0 if

$$\frac{f_A(x_1) f_A(x_2) \dots f_A(x_m)}{f_0(x_1) f_0(x_2) \dots f_0(x_m)} \leq B \quad (4.124)$$

3. Stop sampling and reject H_0 (accept H_A) if

$$\frac{f_A(x_1) f_A(x_2) \dots f_A(x_m)}{f_0(x_1) f_0(x_2) \dots f_0(x_m)} \geq A \quad (4.125)$$

The values of A and B are chosen so that the probabilities of Type I and Type II errors are equal to specified values α and β , respectively.† Exact values of A and B are difficult to obtain. For the small values of α and β typically used in practice, however, good approximations to A and B are given by

$$A \cong \frac{1 - \beta}{\alpha} \quad (4.126)$$

and

$$B \cong \frac{\beta}{1 - \alpha} \quad (4.127)$$

For this sequential test, the sample size required to reach a decision to either accept or reject H_0 is a random variable. It is shown by Wald (1947), however, that, regardless of whether H_0 or H_A is true, the expected value of this sample size, also called the **average sample number** or **ASN**, is always smaller than the fixed sample size required to test the same hypotheses with the same values of α and β using one of the fixed sample size methods of the previous sections. The savings can be substantial. For example, Wald (1947) demonstrates that the average sample

†The constants associated with the SPRT are denoted by A and B , $B < A$, to conform with standard notation. The upper limit A should not be confused with the subscript A in H_A , the alternative hypothesis.

number for the SPRT is about half of the fixed sample size required to test hypotheses about the mean of a normal distribution.

As an illustration, assume that a random variable X has a $N(\mu, \sigma^2)$ distribution with μ unknown but σ^2 known. Suppose that we wish to test the hypothesis

$$H_0: \mu = \mu_0$$

against the alternative

(4.128)

$$H_A: \mu = \mu_A > \mu_0$$

at the α level of significance with specified values of μ_0 , μ_A , and β . From Equation 2.33,

$$f_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-1/2\left(\frac{x - \mu_0}{\sigma}\right)^2\right]$$

and

$$f_A(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-1/2\left(\frac{x - \mu_A}{\sigma}\right)^2\right]$$

After the first observation x_1 is drawn, the ratio of the pdf's is

$$\frac{f_A(x_1)}{f_0(x_1)} = \exp\left[\left(\frac{\mu_A - \mu_0}{\sigma^2}\right)x_1 + \frac{\mu_0^2 - \mu_A^2}{2\sigma^2}\right]$$

and after the m^{th} observation,

$$\frac{f_A(x_1)f_A(x_2)\dots f_A(x_m)}{f_0(x_1)f_0(x_2)\dots f_0(x_m)} = \exp\left[\left(\frac{\mu_A - \mu_0}{\sigma^2}\right)\sum_{i=1}^m x_i + m\left(\frac{\mu_0^2 - \mu_A^2}{2\sigma^2}\right)\right]$$

Referring to Equations 4.123, 4.126, and 4.127, sampling continues as long as

$$\frac{\beta}{1-\alpha} < \exp \left[\left(\frac{\mu_A - \mu_0}{\sigma^2} \right) \sum_{i=1}^m x_i + m \left(\frac{\mu_0^2 - \mu_A^2}{2\sigma^2} \right) \right] < \frac{1-\beta}{\alpha}$$

By taking logarithms and rearranging terms, this becomes

$$\begin{aligned} \left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{\beta}{1-\alpha} \right) + m \left(\frac{\mu_A + \mu_0}{2} \right) \\ < \sum_{i=1}^m x_i < \left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{1-\beta}{\alpha} \right) + m \left(\frac{\mu_A + \mu_0}{2} \right) \end{aligned} \quad (4.129)$$

This inequality can be displayed in a graph with $T_m = \sum_{i=1}^m x_i$ on the vertical axis and m on the horizontal axis. The lines given by

$$U_m = \left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{1-\beta}{\alpha} \right) + m \left(\frac{\mu_A + \mu_0}{2} \right) \quad (4.130)$$

and

$$L_m = \left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{\beta}{1-\alpha} \right) + m \left(\frac{\mu_A + \mu_0}{2} \right) \quad (4.131)$$

are plotted on the graph and define the decision regions for the sequential test. Then, as the observations are taken, the points (m, T_m) are plotted, and sampling continues as long as the sequence of points (m, T_m) falls between the two lines. If a point falls below L_m , the decision is made to stop sampling and accept H_0 . If a point falls above U_m , the decision is made to stop sampling and reject H_0 (i.e., accept H_A). This is illustrated in Figure 4.4.

If a new variable $y_i = x_i - (\mu_0 + \mu_A)/2$ is defined, then the inequality in Equation 4.129 is expressed in terms of y_i as

$$\left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{\beta}{1-\alpha} \right) < \sum_{i=1}^m y_i < \left(\frac{\sigma^2}{\mu_A - \mu_0} \right) \ln \left(\frac{1-\beta}{\alpha} \right) \quad (4.132)$$

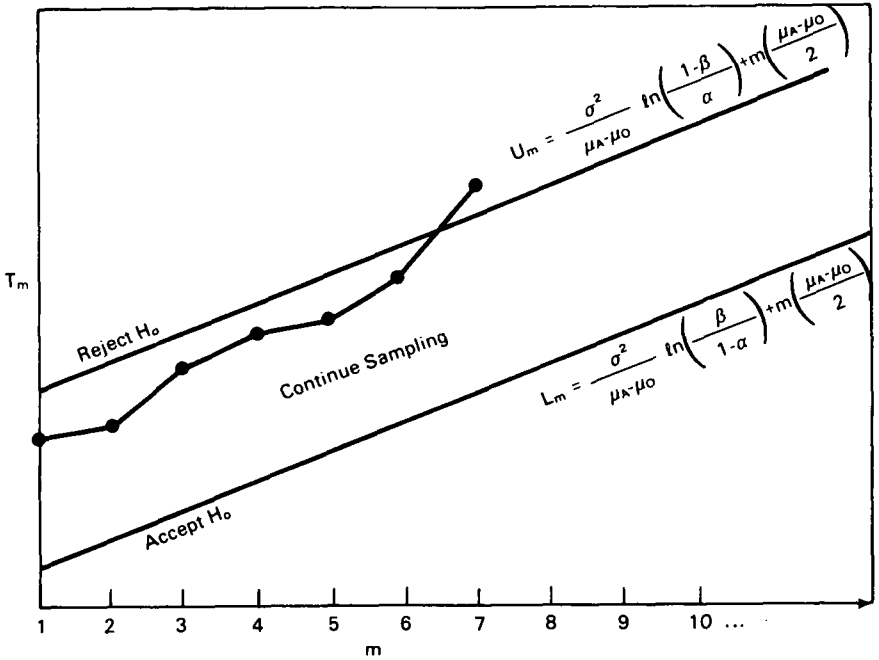


Figure 4.4 Decision regions for the sequential test showing a sampling path leading to the rejection of $H_0: \mu = \mu_0$ in favor of $H_A: \mu = \mu_A > \mu_0$.

The sum $\sum y_i$ is then plotted in a horizontal manner analogous to a cumulative sum chart (introduced in Section 4.7.3).

The one-sided sequential test is easily adapted for testing the two-sided hypothesis

$$H_0: \mu = \mu_0$$

against the alternative

$$(4.133)$$

$$H_A: \mu = \mu_A = \mu_0 + \delta$$

where $\delta = \pm \delta_0$, with specified values of α and β . As with the fixed sample size tests, it is common practice to split α equally across both alternative hypotheses $\mu_A < \mu_0$ and $\mu_A > \mu_0$. The two-sided sequential testing scheme is then equivalent to the two one-sided schemes with $\alpha/2$ and β . If we define a variable $y_i = x_i - \mu_0$, the equations for both the upper and lower pairs of boundary lines in the sequential plot are given by

$$\left(\frac{\sigma^2}{\delta} \right) \ln \left(\frac{\beta}{1 - \alpha/2} \right) + \frac{\delta}{2} m < \sum_{i=1}^m y_i < \left(\frac{\sigma^2}{\delta} \right) \ln \left(\frac{1-\beta}{\alpha/2} \right) + \frac{\delta}{2} m \quad (4.134)$$

where $\delta = \delta_0$ defines one pair of boundary lines, and $\delta = -\delta_0$ defines the other pair. The resulting decision regions are illustrated in Figure 4.5.

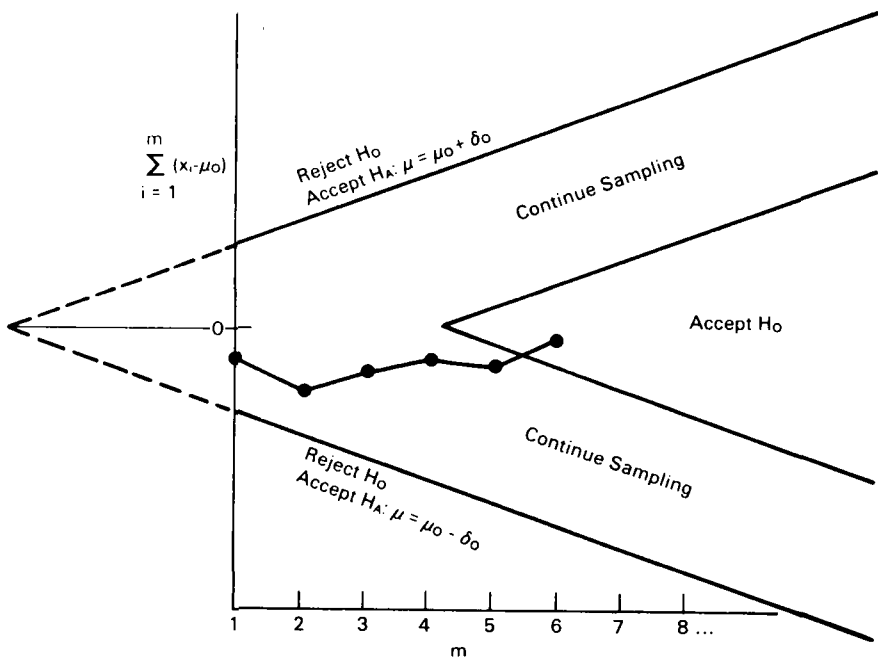


Figure 4.5 Decision regions for the two-sided sequential test showing a sampling path leading to acceptance of H_0 : $\mu = \mu_0$.

Sequential techniques are also useful in applications of attribute sampling (considered in Chapter 17), where the items being sampled are classified as either “success” or “failure.” In Example 4.16, a binomial distribution is assumed where p is the proportion of containers from the process that have weight discrepancies. The plant manager wishes to test the hypothesis

$$H_0: p = 0.05$$

against the alternative

$$H_A: p = 0.09$$

with $\alpha = 0.10$ and $\beta = 0.20$. The results show that a sample size of $n = 175$ is required to ensure these specifications when the test procedure

of Section 4.5.1 is used. By using sequential testing, it is possible to meet the above requirements with an average sample number of fewer than 175 observations. The test proceeds as follows.

Let X be a random variable that takes on the value 1 if the i^{th} container has a weight discrepancy and the value 0 if it does not. Note that X has a binomial distribution with $n = 1$. Referring to Equations 2.20 and 4.122, $f_0(x) = (0.05)^x(0.95)^{1-x}$ and $f_A(x) = (0.09)^x(0.91)^{1-x}$. Let x_i denote the observed value of X for the i^{th} canister. If $T_m = \sum x_i$ is the total number of discrepancies observed after the m^{th} observation, then

$$\frac{f_A(x_1) f_A(x_2) \dots f_A(x_m)}{f_0(x_1) f_0(x_2) \dots f_0(x_m)} = \frac{(0.09)^{T_m} (0.91)^{m-T_m}}{(0.05)^{T_m} (0.95)^{m-T_m}}$$

Then by Equations 4.123, 4.126, and 4.127, sampling continues as long as

$$\left(\frac{0.20}{0.90} \right) < \frac{(0.09)^{T_m} (0.91)^{m-T_m}}{(0.05)^{T_m} (0.95)^{m-T_m}} < \left(\frac{0.80}{0.10} \right)$$

By taking logarithms and rearranging, sampling continues as long as

$$-2.3844 + 0.0682m < T_m < 3.2965 + 0.0682m$$

That is, the decision regions are defined by the lines

$$U_m = 3.2965 + 0.0682m$$

and

$$L_m = -2.3844 + 0.0682m$$

(4.135)

Sampling continues as long as the sequence of points (m, T_m) falls between L_m and U_m . If a point falls below L_m , sampling stops and H_0 is accepted. If a point falls above U_m , sampling stops and H_0 is rejected (H_A is accepted).

From Equation 4.135, $L_m = 0$ when $m = 34.96 \cong 35$. This means that if no weight discrepancies are observed in the first 35 observations, sampling stops, and H_0 is accepted. Thus, if $p = 0$, the process will produce no weight discrepancies in a given period, and only 35 observations are required to reach the decision to accept H_0 . Using results by Wald (1947), if $p = 0.05$, the average sample number, or ASN, required to

reach a decision is 100. If $p = 0.09$, the ASN is 99. The ASN reaches a maximum value of 124 when $p = 0.0682$, that is, when p is equal to the common slope of the two lines in Equation 4.135. Finally, if $p = 1$, the ASN is 4. These five points can be used to plot a crude ASN curve as illustrated in Figure 4.6. Comparing these ASN values with the sample size $n = 175$, required for the testing procedure of Section 4.5.1, gives the reader an appreciation for the potential savings in sampling costs that can be realized by implementing sequential procedures.

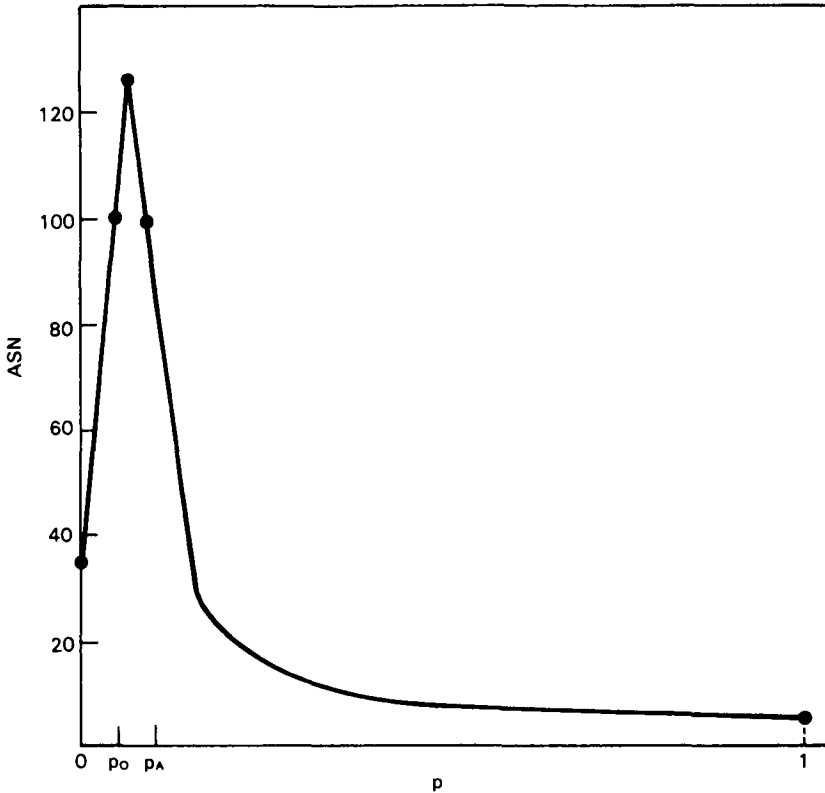


Figure 4.6 ASN curve for testing $H_0: p = 0.05$ vs. $H_A: p = 0.09$, with $\alpha = 0.10$ and $\beta = 0.20$.

Sequential testing schemes for a variety of specific hypotheses are given by Wald (1947), Davies (1956), Duncan (1974), and Wetherill (1975). These references include formulas for constructing OC curves and ASN curves to aid in evaluating sequential testing schemes.

4.7 CONTROL CHARTS

In any manufacturing process, no matter how refined, almost any measurement that can be made on an individual item from the process will vary from item to item. When this variability is confined to chance variation only, the process is said to be in a **state of statistical control**. The process variability might also be affected by one or more assignable causes, however, such as a faulty machine or instrument setting, worn components, an unskilled operator, unsatisfactory raw materials, and so on. Such assignable causes of variation usually have an adverse effect on product quality. Thus, it is important to detect serious departures from a state of statistical control as soon as possible after they occur. **Control charts** are widely used for this purpose.

Whereas many practitioners associate control charts with statistical quality control (SQC) or statistical process control (SPC), the concept of quality control involves a management philosophy that goes well beyond simply monitoring a process or system. In fact, the control chart methods presented in this section are only one small set of the many statistical and nonstatistical techniques that are useful in SQC and SPC applications. Thus, this section is not intended to be a guide for quality control methodology. Instead, its purpose is to introduce a few techniques that have proven to be quite useful in many SQC and SPC applications. For detailed treatments of quality control philosophies, concepts, and techniques, the reader is referred to such texts as Deming (1982), Ishikawa (1976), Juran and Gryna (1980), and Wadsworth et al. (1986).

It should also be noted that the techniques presented in this section can be programmed on a computer. In fact, many commercially available computer software packages will perform these techniques.

A control chart consists of a **center line (CL)**, corresponding to the average quality at which the process should perform when in a state of statistical control, and two **control limits**, called the **upper control limit (UCL)** and the **lower control limit (LCL)**. A typical control chart is illustrated in Figure 4.7. The control limits are chosen such that values falling between them can be attributed to random chance, but values falling outside them indicate a lack of statistical control. The general approach is to periodically take a random sample from the process, compute an appropriate quantity (such as the sample mean or total), and plot this quantity on the control chart. When a sample value falls outside the control limits, an investigation into possible assignable causes of variation is undertaken in an attempt to identify and resolve the cause of this unusual outcome. Even when all sample values fall between the control limits, however, a trend or some other nonrandom pattern may indicate that some action is necessary to avoid a more serious problem.

One use of control charts is to monitor an important measurement on a succession of mass-produced objects from a process that has been previ-

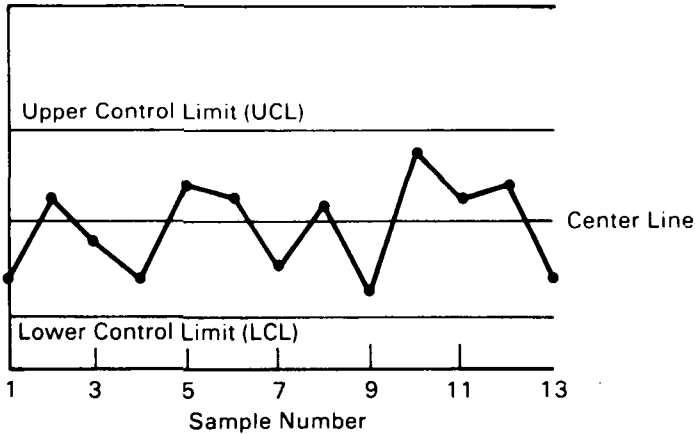


Figure 4.7 A typical control chart.

ously brought into a state of statistical control. In such situations, there has usually been an evolution of the process to the point where the distribution of the measurements is approximately known, or at least its mean and variance are known. Control charts used for this purpose are called **theoretical control charts**. They provide a means for detecting departures from an existing state of statistical control.

Another major use of control charts is to provide guidance in establishing a state of statistical control. Control charts used for this purpose depend upon computations made from a historical sequence of samples of measurements and are called **empirical control charts**.

Both theoretical and empirical control charts for some applications are presented in this section with formulas and procedures for their implementation. For more detailed presentations of these and other control chart applications, see Duncan (1974) or Burr (1976).

4.7.1 Theoretical Control Charts

When dealing with a quality characteristic, it is customary to exercise control over both the mean and variability of the characteristic. In this section, three basic types of theoretical control charts are presented. The first type is quite general and is typically used to monitor means, proportions, and rates. The other two types, based upon the sample variance and the sample range, are used to monitor variability.

4.7.1.1 Theoretical Control Charts for Means, Proportions, and Rates

Consider a manufacturing process which is in a state of statistical control with respect to a particular quality characteristic. It is of interest to construct a control chart for the characteristic so that if the process is

thrown out of control with respect to the characteristic, the control chart will detect this event. Specifically, suppose that a process is in a state of statistical control and that a measurement is made on each object in a random sample of size n taken from a particular part of the production line at a given time. If a statistic Y is computed from the sample (for example, the sample mean \bar{X} or the sample total), then in successive samples, the values of Y , say Y_1, Y_2, \dots , will behave as independent random variables from a common distribution. This common distribution has a mean μ_Y and a standard deviation σ_Y , which are known if the distribution of the measurements on the successive objects manufactured under the state of statistical control is known.

It must be emphasized that all of the control chart methods presented in Section 4.7 require that the successive Y_i items are independent and that each Y_i is based upon a random sample. If the Y_i items are correlated, application of these control chart techniques can lead to erroneous conclusions. For example, a misuse of control charts in nuclear material accounting would be for monitoring successive inventory differences, which are definitely correlated. The assumption that each Y_i is based upon a random sample requires that the observations within each sample be uncorrelated. When this assumption is in doubt, specialized techniques such as those presented by Alt et al. (1977) and Alt and Deutsch (1979) should be used.

In many situations, the distribution of Y can be approximated by a normal distribution with mean μ_Y and standard σ_Y . In such cases, the probability that Y will fall in any given interval when the process is under statistical control is easily computed. For example, the probability is 0.9973 that Y will fall inside the limits $\mu_Y \pm 3\sigma_Y$. Of course, the validity of such statements depends upon how closely the distribution of Y is approximated by the $N(\mu_Y, \sigma_Y^2)$ distribution.

A **three-sigma control chart**, also known as a **Shewhart chart**, is set up by drawing three horizontal lines:

$$\begin{aligned} \text{UCL} &= \mu_Y + 3\sigma_Y \\ \text{CL} &= \mu_Y \\ \text{LCL} &= \mu_Y - 3\sigma_Y \end{aligned} \tag{4.136}$$

The successive values Y_1, Y_2, \dots , are then plotted against the sample number on this chart. This is visualized by referring to Figure 4.7.

Suppose that something causes the distribution of Y to shift by an amount δ , so that Y now has mean $\mu_Y + \delta$ and standard deviation σ_Y . If Y is approximately normally distributed, then the successive values Y_1, Y_2, \dots , will behave like a random sample from the $N(\mu_Y + \delta, \sigma_Y^2)$ distribu-

tion. The probability that a given Y_i will fall outside the three-sigma control limits is given by

$$1 - [\Phi(3 - \delta/\sigma_Y) - \Phi(-3 - \delta/\sigma_Y)] \quad (4.137)$$

where $\Phi(z)$ is obtained from Table A3. Thus, the larger the absolute value of δ/σ_Y , the higher the probability that the control chart will detect the shift in the distribution of Y .

The variable Y is used in the previous discussion to represent several possible statistics. Some of the statistics frequently used in practice are the sample mean \bar{X} , the number of defective items X in a sample of size n , the sample proportion defective $\hat{p} = X/n$, and the count C of the number of events per unit time (or space, etc.). The specifications for three-sigma control charts for each of these statistics are given in Table 4.2. In examining Table 4.2, note that the normal distribution is presented in Section 2.5.2, the binomial in Section 2.4.2, and the Poisson in Section 2.4.3.

TABLE 4.2

Three-Sigma Control Chart Specifications for Four Statistics

Statistic Y	Distribution	UCL	CL	LCL
Sample mean: \bar{X}	Normal	$\mu + 3\sigma/\sqrt{n}$	μ	$\mu - 3\sigma/\sqrt{n}$
Sample number defective: X	Binomial	$np + 3\sqrt{np(1-p)}$	np	$np - 3\sqrt{np(1-p)}$
Sample proportion defective: \hat{p}	Binomial	$p + 3\sqrt{p(1-p)/n}$	p	$p - 3\sqrt{p(1-p)/n}$
Total count: C	Poisson	$\mu + 3\sqrt{\mu}$	μ	$\mu - 3\sqrt{\mu}$

Control charts employing these three-sigma limits are called **theoretical control charts** since the CL, UCL, and LCL are determined by using known parameter values from the distributions of the measurements.

Although in many practical situations the statistic Y is approximately normally distributed for a process under statistical control, even when Y is not normally distributed, it is customary in practice to establish a three-sigma control chart as defined above for Y . In such cases, for a process under statistical control, the probability that Y will fall between the UCL and LCL is not 0.9973. Nevertheless, the UCL and LCL are referred to as three-sigma control limits, where the probability is quite small that a given Y_i will fall outside the limits when the process is in a state of statistical control. If the distribution of Y is known (or approximately known), this probability can be computed (or approximated). The following discussion demonstrates that this probability will never exceed 0.11 and will be much smaller in most applications.

A basic theorem in probability theory, known as Chebyshev's inequality, guarantees that if Y has a finite mean μ_Y and variance σ_Y^2 , then no matter how Y is distributed,

$$\Pr(\mu_Y - k\sigma_Y < Y < \mu_Y + k\sigma_Y) \geq 1 - \frac{1}{k^2} \quad (4.138)$$

That is, for any distribution, the probability that Y will fall outside the interval $\mu_Y \pm k\sigma_Y$ is at most $1/k^2$. Thus, for a three-sigma control chart, the probability that a given Y_i will fall outside the control limits is at most $1/9 = 0.11$ when the process is under statistical control and is near 0.0027 if Y is approximately normally distributed. For most distributions encountered in practice, Chebyshev's inequality is overly conservative for estimating this probability (it overestimates the true value). Thus, it would rarely, if ever, be used in practice, and it is used here only to demonstrate that the probability will never exceed 0.11 for three-sigma control limits. See Rohatgi (1984) for a detailed presentation of Chebyshev's inequality.

When Y is from either the binomial or the Poisson distribution, and the normal approximation is questionable, the concepts from Sections 4.5.1 and 4.5.4 on exact tests are used to determine control limits based upon the true distribution of Y . This involves finding upper and lower control limits such that when a state of statistical control exists, $\Pr(\text{LCL} < Y < \text{UCL}) \cong 1 - \alpha$. To be analogous to a three-sigma control chart for a normally distributed variable, $1 - \alpha = 0.9973$ and the LCL and UCL are determined such that $\Pr(Y \leq \text{LCL}) \cong \alpha/2$ and $\Pr(Y \geq \text{UCL}) \cong \alpha/2$ where $\alpha/2 = 0.00135$.

Example 4.21 In this example, given by Jaech (1973), a theoretical control chart is constructed for the average percent uranium in successive batches of UO_2 powder. When the process is in a state of statistical control, it is assumed that $\mu = 87.60\%$ and $\sigma = 0.06\%$. Thus, a three-sigma control chart is defined by the following quantities:

$$\text{UCL} = 87.60 + 3(0.06) = 87.78$$

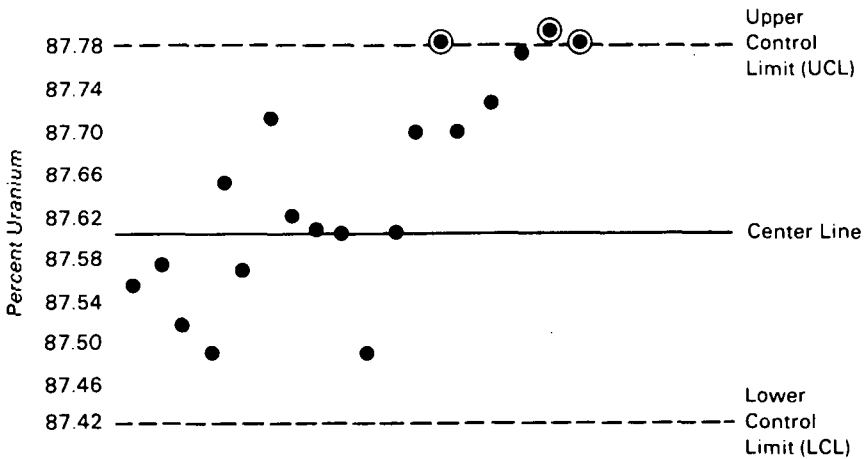
$$\text{CL} = 87.60$$

$$\text{LCL} = 87.60 - 3(0.06) = 87.42$$

The data from 19 successive batches are given in the order observed.

87.54	87.56	87.47	87.72
87.56	87.71	87.60	87.77
87.50	87.61	87.69	87.79
87.47	87.60	87.78	87.78
87.64	87.60	87.69	

These values are plotted on the control chart below, where the circled points are greater than or equal to the UCL.



The batches with circled values may have true means larger than 87.60%. In fact, since the last seven points are all above the CL, it appears likely that there has been an upward shift in the process mean.

4.7.1.2 Theoretical Control Charts for the Variance

Let σ^2 be the known variance of the measurements on objects manufactured by a process which is in a state of statistical control. Suppose that we wish to establish a control chart to monitor σ^2 based upon the sample variances, S_1^2, S_2^2, \dots , from a succession of independent samples of size n from a normal distribution. In practice it is customary to use only a UCL for monitoring σ_i^2 . If the probability is to be α that any S_i^2 would fall above the UCL, then by analogy with the hypothesis test defined by Equations 4.21, 4.23, and 4.24, the UCL is the horizontal line given by

$$\text{UCL} = \frac{\sigma^2 \chi_{1-\alpha}^2(n-1)}{n-1} \quad (4.139)$$

To work in terms of S_i rather than S_i^2 , take the square root of the right-hand side of Equation 4.139.

In some applications, it is also of interest to detect unusually small values of S_i^2 . In such cases, both upper and lower control limits are needed. These are given by

$$UCL = \frac{\sigma^2 \chi_{1-\alpha/2}^2(n-1)}{n-1}$$

and (4.140)

$$LCL = \frac{\sigma^2 \chi_{\alpha/2}^2(n-1)}{n-1}$$

To work in terms of S_i rather than S_i^2 , the UCL and LCL are determined as the square roots of the right-hand sides of Equations 4.140.

4.7.1.3 Theoretical Control Charts for the Range

Another technique for monitoring the variance is to set up a control chart based upon the sample range R . Since R is simply the difference between the maximum and minimum values observed in a sample, it is easier to compute than S^2 . Intuitively, if the process variance σ^2 increases, R is also expected to increase. If the process is in a state of statistical control and if σ^2 is known, a three-sigma theoretical control chart for R has control limits

$$LCL = D_1\sigma$$

and (4.141)

$$UCL = D_2\sigma$$

where D_1 and D_2 depend on n . For samples from a normal distribution, D_1 and D_2 are tabulated in Table 4.3 for $n = 2, 3, \dots, 10$. In many applications, n is smaller than 10. For guidance on how to proceed with $n > 10$, see Duncan (1974) or Burr (1976).

As with the control chart for the variance, the LCL is rarely used in practice. In fact, when $n < 7$, Table 4.3 gives $D_1 = 0$, which is equivalent to no LCL. This is because the sample range will never be less than zero, but the lower three-sigma control limit would be negative (and thus meaningless) when $n < 7$; thus, no LCL is used.

TABLE 4.3
Values of D_1 and D_2 for
Three-Sigma Theoretical
Control Charts for the Range

n	D_1	D_2
2	0	3.686
3	0	4.358
4	0	4.698
5	0	4.918
6	0	5.078
7	0.205	5.203
8	0.387	5.307
9	0.546	5.394
10	0.687	5.469

4.7.2 Empirical Control Charts

In some situations a manufacturing process is under way, but a satisfactory state of statistical control has not yet been established. Thus, the measurements taken on successive manufactured objects do not behave like a random sample from a distribution with the desired mean and variance. If assignable causes of this lack of statistical control can be identified and resolved, however, then the behavior of the measurements should more closely resemble a random sample from the desired distribution.

When the parameters of the measurement distribution are unknown, they must be estimated using historical data from the process. Thus, reaching a state of statistical control by identifying and resolving assignable causes of extraneous variability is somewhat iterative and evolutionary. Briefly speaking, historical data from the process are used to estimate parameters and to construct **empirical control charts** on which past and ongoing process data are plotted. From these plots, one or more of the assignable causes of variability might be detected, identified, and resolved. The parameter estimates and control charts are then refined by removing obvious outlying points from the data, and the above steps are repeated. This usually continues over time until the process is refined to the point where it is in a satisfactory state of statistical control and the parameters of the measurement distribution are assumed to be known. Once this has been achieved, the theoretical control chart techniques of the previous section are implemented.

This section presents empirical control chart techniques for means, proportions, rates, and variances.

4.7.2.1 Empirical Control Charts for Means, Proportions, and Rates

Consider first a control chart for the mean of a normally distributed process, when μ and σ^2 are unknown. Assume that k independent random samples, each of size n , have been taken from the process, such that the objects within each sample are closely spaced in time, but the samples themselves have been taken periodically and may or may not be closely spaced in time. Let $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ be the sample means and S_1, S_2, \dots, S_k be the sample standard deviations of the k samples. Proceeding in this manner helps to ensure that the sample standard deviations, S_i , reflect only the random variability in the process and that they are not inflated by shifts in the process mean μ . It also helps to ensure that shifts in μ are reflected by the sample means \bar{X}_i .

If the process is in a state of statistical control, the sample means, \bar{X}_i , will follow the $N(\mu, \sigma^2/n)$ distribution. An estimate of μ that is used in practice is the average of the k sample means, given by

$$\bar{\bar{X}} = \sum_{i=1}^k \bar{X}_i/k \quad (4.142)$$

Similarly, an estimate of σ is obtained by averaging the k sample standard deviations and is given by

$$\bar{S} = \sum_{i=1}^k S_i/k \quad (4.143)$$

The following procedure is based upon the normal distribution but is satisfactory if the sample means \bar{X}_i satisfy the rules of thumb for application of the Central Limit Theorem (Section 2.5.2.1). The UCL, CL, and LCL for a three-sigma empirical control chart for the mean are given as follows:

$$\begin{aligned} \text{UCL} &= \bar{\bar{X}} + A_1 \bar{S} \\ \text{CL} &= \bar{\bar{X}} \\ \text{LCL} &= \bar{\bar{X}} - A_1 \bar{S} \end{aligned} \quad (4.144)$$

where A_1 depends on the sample size n and is tabulated in Table 4.4 for $n = 2, 3, \dots, 10$. A_1 is a quantity such that $A_1 \bar{S}$ is an unbiased estimate of $3\sigma/\sqrt{n}$. For $n > 10$, see Table V in Burr (1976).

An alternate method of estimating $3\sigma/\sqrt{n}$ uses the k sample ranges R_1, R_2, \dots, R_k rather than the sample standard deviations. Thus, it is computationally simpler than the previously mentioned procedure, but it makes

TABLE 4.4
Values of A_1 and A_2 for
Three-Sigma Empirical
Control Charts for the Mean

n	A_1	A_2
2	2.659	1.880
3	1.954	1.023
4	1.628	0.729
5	1.427	0.577
6	1.287	0.483
7	1.182	0.419
8	1.099	0.373
9	1.032	0.337
10	0.975	0.308

less efficient use of the sample information. Let \bar{R} be the average of the k sample ranges, given by

$$\bar{R} = \sum_{i=1}^k R_i/k \quad (4.145)$$

The UCL, CL, and LCL for a three-sigma empirical control chart for the mean are given as follows:

$$\begin{aligned} \text{UCL} &= \bar{\bar{X}} + A_2\bar{R} \\ \text{CL} &= \bar{\bar{X}} \\ \text{LCL} &= \bar{\bar{X}} - A_2\bar{R} \end{aligned} \quad (4.146)$$

where A_2 depends on the sample size n and is tabulated in Table 4.4 for $n = 2, 3, \dots, 10$. A_2 is a quantity such that $A_2\bar{R}$ is an unbiased estimate of $3\sigma/\sqrt{n}$. Duncan (1974) and Burr (1976) provide guidance on how to proceed with $n > 10$.

When preliminary samples are used to estimate limits for control charts, it is customary to treat these limits as trial values. Therefore, the k sample means are plotted on the chart, and any points that fall outside the control limits are investigated. If assignable causes for these points are discovered, they are eliminated from the data. Then $\bar{\bar{X}}$ and either $\bar{\bar{S}}$ or \bar{R} are recomputed, and the values of the UCL, CL, and LCL are updated. Proceeding in this manner, a state of statistical control may eventually be established.

Example 4.22 In an automated process, UO_2 powder containers are filled, weighed, sealed, and labeled. This process was recently set up by a facility, and a state of statistical control has not yet been established. Five containers are randomly selected from the process each day. The sampled containers are weighed, and their weights recorded in kilograms. The sample means and ranges for the first 20 days are given below.

Day	\bar{X}_i	R_i	Day	\bar{X}_i	R_i
1	22.316	0.04	11	22.298	0.04
2	22.330	0.03	12	22.340	0.04
3	22.350	0.04	13	22.330	0.10
4	22.322	0.04	14	22.348	0.04
5	22.338	0.02	15	22.356	0.07
6	22.384	0.03	16	22.308	0.06
7	22.316	0.04	17	22.330	0.05
8	22.368	0.10	18	22.316	0.03
9	22.350	0.15	19	22.282	0.09
10	22.340	0.06	20	22.338	0.06

$$\bar{\bar{X}} = 22.333 \quad \bar{R} = 0.0565$$

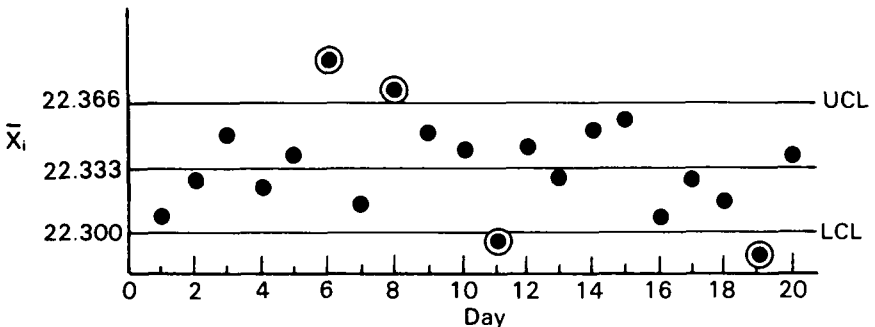
Entering Table 4.4 with $n = 5$, the value $A_2 = 0.577$ is obtained. An empirical control chart for the mean weight has limits given by

$$\text{UCL} = 22.333 + (0.577)(0.0565) = 22.366$$

$$\text{CL} = 22.333$$

$$\text{LCL} = 22.333 - (0.577)(0.0565) = 22.300$$

The sample means are plotted on the control chart below, where the means for days 6, 8, 11, and 19 fall outside the control limits.



Assume that the cause of these outlying points has been identified and resolved. If days 6, 8, 11, and 19 are removed from the data, the recomputed quantities are

$$\begin{aligned}\bar{X} &= 22.333 \\ \bar{R} &= 0.0544 \\ \text{UCL} &= 22.364 \\ \text{CL} &= 22.333 \\ \text{LCL} &= 22.302\end{aligned}$$

Note that CL has not changed and that the UCL and LCL have changed only slightly. By examining the previous control chart, it is clear that removing the sample means for days 6 and 8 balances the effect of removing those for days 11 and 19. This is coincidental and may not be typical of the changes and refinements that might be observed in practice. All of the 16 remaining sample means fall within these updated limits.

Further refinement of this process based upon variability is discussed in Example 4.23.

Empirical control charts for the proportion of defectives also assume that k independent random samples, each of size n , have been taken periodically from the process. It should be noted, however, that the required sample sizes for such attribute sampling plans are considerably larger than those where measurements, rather than classifications, are made. This is analogous to the required sample size calculations in Examples 4.16 and 4.17.

Let X_i be the number of defective objects observed in the i^{th} random sample of size n . Then $\hat{p}_i = X_i/n$ is the corresponding sample proportion defective. Assume that X_i is a random variable having a binomial distribution with parameter p . A combined estimate of p is obtained by averaging the sample proportions and is given by

$$\bar{p} = \sum_{i=1}^k \hat{p}_i / k = \sum_{i=1}^k \frac{X_i}{nk} \quad (4.147)$$

Thus, a three-sigma empirical control chart for p is defined by the values

$$\begin{aligned}\text{UCL} &= \bar{p} + 3\sqrt{\bar{p}(1 - \bar{p})/n} \\ \text{CL} &= \bar{p}\end{aligned} \quad (4.148)$$

$$\text{LCL} = \bar{p} - 3\sqrt{\bar{p}(1 - \bar{p})/n}$$

If p is small, the lower control limit may be negative. When this happens, the LCL is replaced by zero. These control limits are based upon the normal approximation to the binomial (Section 2.5.2.2). When the rules of thumb for applying this approximation are not satisfied, it is best to use control limits determined directly from the binomial distribution by analogy with the exact methods described in Section 4.5.1.

In some applications, count rates are monitored. For example, nondestructive assay (NDA) instruments may be checked daily by analyzing a standard. The response obtained from an instrument is typically counts per minute (or counts per some other unit of time). When the instrument is functioning properly and is not in need of adjustment or repair, the count rate should follow a Poisson distribution with parameter μ . If μ is known, a theoretical control chart as indicated in Table 4.2 is used to monitor the daily check readings from an instrument. If μ is unknown, then an empirical control chart can be set up by the following general procedure.

Let C be the total counts observed in a time interval of specified length. If k such counts are available from a process of interest, let C_i be the count from the i^{th} time interval. An estimate of μ is given by

$$\bar{C} = \sum_{i=1}^k C_i/k \quad (4.149)$$

A three-sigma empirical control chart for the daily counts is defined by the values

$$\begin{aligned} \text{UCL} &= \bar{C} + 3\sqrt{\bar{C}} \\ \text{CL} &= \bar{C} \\ \text{LCL} &= \bar{C} - 3\sqrt{\bar{C}} \end{aligned} \quad (4.150)$$

In the NDA instrument example, the historical daily check counts are plotted on the control chart. If the count for any of the days falls outside the control limits, an investigation into the cause is conducted. If assignable causes are found and resolved for any of these outlying counts, they are removed from the data and the control limits are recomputed. Thereafter, when a count falls outside the control limits, the instrument is examined, and any necessary repairs or adjustments are made.

Note that the control limits in Equation 4.150 are based upon the normal approximation to the Poisson discussed in Section 4.5.5. When

$\bar{C} < 50$, control limits should be determined directly from the Poisson distribution by analogy with the exact methods described in Section 4.5.4.

4.7.2.2 Empirical Control Charts for the Range

When it is of interest to monitor the variance σ^2 , but σ^2 is unknown, an empirical control chart based upon the sample range can be used. Assume that k independent random samples of size n have been taken from the process of interest. It must also be assumed that the individual observations within each sample are from a normal distribution. Then \bar{R} , given by Equation 4.145, is used to set up an empirical control chart for the range with control limits given by

$$\begin{aligned} \text{LCL} &= D_3 \bar{R} \\ \text{UCL} &= D_4 \bar{R} \end{aligned} \quad (4.151)$$

where D_3 and D_4 depend on the sample size n and are tabulated in Table 4.5 for $n = 2, 3, \dots, 10$. Duncan (1974) and Burr (1976) give guidance on how to proceed when $n > 10$.

TABLE 4.5
Values of D_3 and D_4 for Three-Sigma
Empirical Control Charts for the Range

n	D_3	D_4
2	0	3.267
3	0	2.575
4	0	2.282
5	0	2.115
6	0	2.004
7	0.076	1.924
8	0.136	1.864
9	0.184	1.816
10	0.223	1.777

As with the charts for the variance, the LCL is seldom of interest. In fact, when $n < 7$, Table 4.5 gives $D_3 = 0$, which is equivalent to no LCL. This is because the sample range will never be less than zero, but the lower three-sigma control limit would be negative (and thus meaningless) when $n < 7$; thus, no LCL is used.

Example 4.23 In Example 4.22, after removing days 6, 8, 11, and 19 from the data, the value $\bar{R} = 0.0544$ is computed. Entering Table 4.4 with $n = 5$, the values $D_3 = 0$ and $D_4 = 2.115$ are obtained. Thus, an empirical control chart for the range has control limits given by

$$\text{LCL} = 0$$

$$\text{UCL} = (2.115)(0.0544) = 0.115$$

Without plotting the 16 sample ranges, note that day 9 has $R = 0.15 > 0.115$.

Suppose that the cause of this large value has been resolved. If day 9 is removed from the data, the recomputed values are

$$\bar{R} = 0.0480$$

$$\text{LCL} = 0$$

$$\text{UCL} = 0.102$$

and all 15 remaining sample ranges are within these new limits. After removing day 9, the control chart for the mean should also be updated.

Note that day 13 has $R = 0.10$, which is just inside the UCL. It would be advisable to investigate the cause of this value. If it is removed from the data, then both the control charts for the range and for the mean should be updated.

In any case, future values of \bar{X}_i and R_i will be incorporated and the control limits for both the mean and range charts refined until a satisfactory state of statistical control is established.

4.7.3 Cumulative Sum Charts

The theoretical and empirical control charts discussed so far control the probability of a Type I error. Thus, as long as the process is in control, only rarely will a plotted point fall outside the three-sigma control limits. In practice, however, the Type II error can have serious consequences. That is, failing to detect when a process is out of control can adversely affect product quality. The **cumulative sum chart** (CUSUM chart), together with the **V-mask**, which is an adaptation of the sequential procedure discussed in Section 4.6, is a charting technique which controls the probabilities of both the Type I and Type II errors. An excellent survey of the use of CUSUM charts is given by Ewan (1963).

Although the cumulative sum chart technique is quite general in scope, it is introduced by considering only the case where the measurements are from a normal distribution and the mean is to be monitored. That is, if the process is in a state of statistical control, then the measurements follow the $N(\mu_0, \sigma^2)$ distribution. It is further assumed that μ_0 and σ are either known or have been estimated from a large number of measurements taken at times when the process was thought to be in control. Techniques from Section 4.7.2.1 can be used on historical process data to develop refined estimates of μ_0 and σ . In this section, the symbols μ_0 and σ are used to denote either the known values or the refined estimates.

Cumulative sum charts can be applied to either a sequence of independent single measurements or to a sequence of sample means computed from independent samples of size n . For convenience, the notation Y_1, Y_2, \dots , is used to denote the sequence of interest, and if the process is in a state of statistical control, the observed values of the Y_i behave like a random sample from the $N(\mu_0, \sigma_Y^2)$ distribution. Note that if Y_i is a single measurement, then $\sigma_Y^2 = \sigma^2$; if $Y_i = \bar{X}_i$, then $\sigma_Y^2 = \sigma^2/n$.

Instead of plotting the sequence Y_1, Y_2, \dots, Y_m , the **cumulative sum**, or **CUSUM**, given by

$$T_m = \sum_{i=1}^m (Y_i - \mu_0) \quad (4.152)$$

is plotted against m . If the process is in a state of statistical control, T_m is expected to fluctuate randomly about 0. If, however, the mean should shift from μ_0 to $\mu_0 + \delta$, then T_m is expected to increase by δ for each new observation if δ is positive and to decrease if δ is negative.

The scale used for plotting a CUSUM chart is of some importance and should be considered in advance. The horizontal scale should be chosen such that consecutive points can be easily distinguished. It is recommended that the relationship between the scales of the horizontal and vertical axes be such that the distance between consecutive observations on the horizontal axis is approximately equal to the distance $2\sigma_Y$ on the vertical axis. This is illustrated in Figure 4.8.

Some examples of data and the corresponding CUSUM charts are illustrated in Figure 4.9. The data in Figure 4.9A fluctuate about μ_0 , and the CUSUM plot is quite flat and fluctuates about 0. In Figure 4.9B, the data are all above μ_0 , and the CUSUM is continually increasing. Similarly, the data in Figure 4.9C are all below μ_0 , and the CUSUM continually decreases.

Now suppose that in a continuing process a shift in the process mean from μ_0 to $\mu_0 + \delta$ occurred which affected all Y_i values from Y_r through Y_m (where $r < m$). Figure 4.10 shows a plot of the data and the

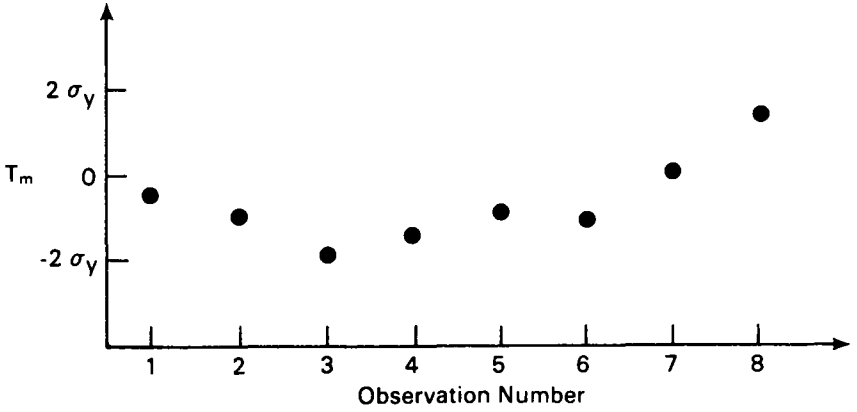
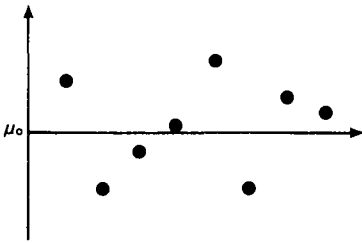
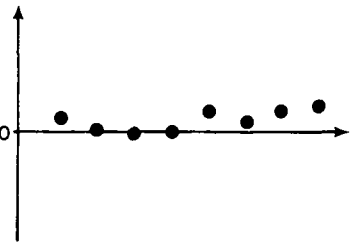


Figure 4.8 Recommended scales for CUSUM charts. $2\sigma_y$ units on the vertical scale should equal the distance between observations on the horizontal axis.

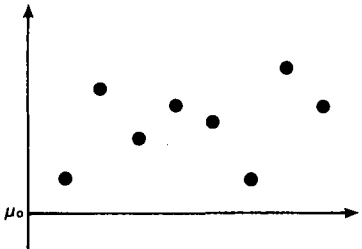
A. Data Fluctuates about μ_0



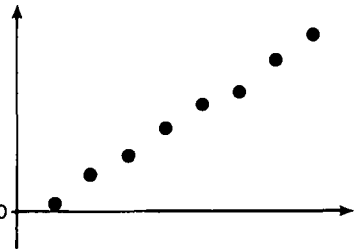
Cusum Chart Essentially Flat



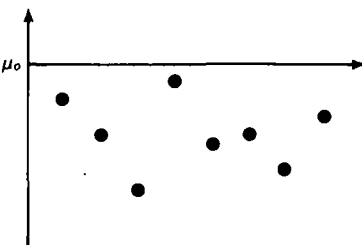
B. Data Above μ_0



Cusum Chart Increasing



C. Data Below μ_0



Cusum Chart Decreasing

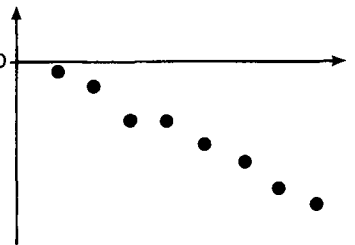


Figure 4.9 Examples of data and associated cumulative sum charts.

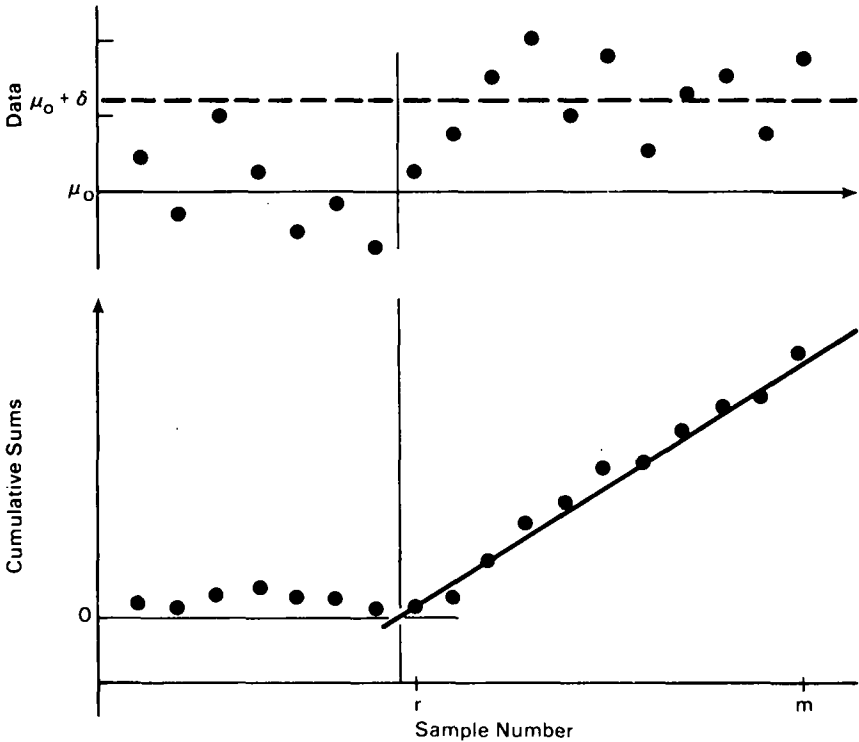


Figure 4.10 Data plot and CUSUM chart showing the effect of a shift in the process mean from μ_0 to $\mu_0 + \delta$.

CUSUM chart for such a situation, where δ is positive. Notice how the CUSUM chart responds quickly to the change, providing a much clearer visual alert than the data plot. Also, a straight line drawn through the 11 most recent points on the CUSUM chart provides a good indication of when the shift occurred. Let T_r be the CUSUM just after the point where this straight line intersects a horizontal line drawn at CUSUM = 0. The point (r, T_r) is called the **turning point** and is the best guess as to when the shift occurred. An estimate of the amount of shift δ is given by

$$\hat{\delta} = \frac{T_m - T_r}{m - r} \quad (4.153)$$

where T_m is the most recent CUSUM value and T_r is defined above.

Although the CUSUM chart alone provides valuable information about the process, it does not provide any criteria for deciding whether an observed change warrants action. Since it is desirable to monitor the process mean subject to specified probabilities of making a Type I or Type II

error, the quantities μ_0 , σ_Y , α , β , and the value of δ that would cause concern must be specified. The hypotheses of interest are

$$H_0: \mu = \mu_0$$

$$H_A: \mu = \mu_0 + \delta$$

with specified values of α and β . Then, by analogy with the two-sided sequential test, a **V-mask** is constructed.

The V-mask, illustrated in Figure 4.11, is identical to the "V" produced by the two-sided sequential testing procedure (Figure 4.5), except that the mask is reversed and placed adjacent to the most recently plotted point

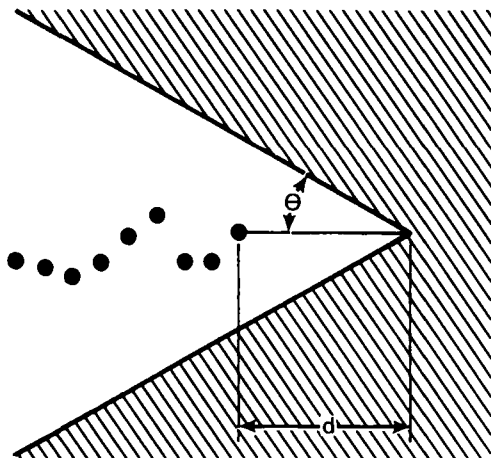


Figure 4.11 Example of a V-mask.

(m, T_m). To construct a V-mask like the one in Figure 4.11, the angle θ is given by

$$\theta = \tan^{-1} \left(\frac{|\delta|}{2C} \right) \quad (4.154)$$

where C is a scale factor giving the number of units on the vertical axis per unit on the horizontal axis. For example, the recommended scales, 2σ vertical units per 1 horizontal unit, gives $C = 2\sigma/1 = 2\sigma$. The **lead distance** d is given by

$$d = \frac{2\sigma^2}{\delta^2} \log \left(\frac{1 - \beta}{\alpha/2} \right) \quad (4.155)$$

The V-mask is placed such that the V is a lead distance d in front of the most recently plotted point. If a previously plotted point falls beyond the "arms" of the V-mask, reject the hypothesis that the process mean is μ_0 .

The application of the V-mask follows the algorithm given below:

1. Prepare the CUSUM chart, taking care to use appropriate scales.
2. Specify δ , α , β , and σ ; determine θ and d ; and draw the appropriate V-mask on a sheet of transparent paper.

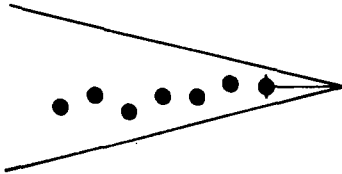
3. Position the V-mask over each successive point on the CUSUM chart, with the middle line of the V-mask parallel to the horizontal axis (Figure 4.11).

4. When the V-mask is positioned over a given point, examine all points to the left of this point with respect to the arms of the V-mask.

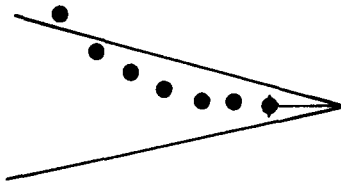
This algorithm can be programmed on a computer. In fact, there are many commercially available computer software packages that will perform a sequential V-mask analysis.

Some of the patterns that might be observed are illustrated and briefly interpreted below:

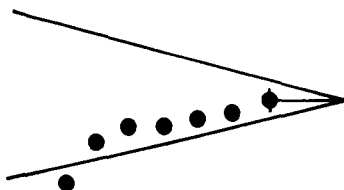
A. All points within the V indicate a stable process around μ_0 .



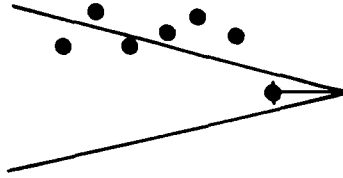
B. A downward trend which crosses the V indicates that the process mean is significantly below μ_0 .



C. An upward trend which crosses the V indicates that the process mean is significantly above μ_0 .



D. A sudden one-point drop indicates an outlying value whose cause should probably be investigated.



In implementing the CUSUM and V-mask technique, the successive observations *must* be independent (uncorrelated). Also, it is advisable to restart the cumulative sum calculations whenever a significant shift is indicated by the V-mask technique. Otherwise, the CUSUM statistic T_m may never return to being near zero, even if the process returns to a state of statistical control and remains in control indefinitely.

The CUSUM chart provides information about the entire history of a process mean. As the number of terms in the cumulative sum becomes large, however, some types of shifts in the mean could be “smoothed over” or balanced out by previous undetected shifts and go unnoticed. Moving sum charts are a compromise between CUSUM charts and control charts. A moving sum, as the name implies, includes only the most recent set of k observations. Then, each time a new observation is added to the sum, the “oldest” observation is dropped from the sum.

If a moving sum is to include only the most recent k observations in a sequence of $m > k$ observations, the formula for the moving sum is

$$T_m(k) = \sum_{i=m-k+1}^m (Y_i - \mu_0) \quad (4.156)$$

When $k = 1$, the moving sum is simply $(Y_m - \mu_0)$, the most recent observation minus μ_0 , and the moving sum chart reduces to a control chart as presented in Section 4.7.1.

Bauer and Hackl (1978, 1980) provide formulas and procedures for a variety of applications of moving sum charts.

REFERENCES

- Alt, F. B., S. J. Deutsch, and J. W. Walker, 1977, Control Charts for Multivariate Correlated Observations, *31st Annual Technical Conference Transactions*, American Society for Quality Control, pp. 360-369.
- , and S. J. Deutsch, 1979, Economic Control Charts for Correlated Observations, *33rd Annual Technical Conference Transactions*, American Society for Quality Control, pp. 743-750.
- Aspin, A. A., 1949, Tables for Use in Comparisons Whose Accuracy Involves Two Variances, Separately Estimated, *Biometrika*, 36: 290-296.

- Bauer, P., and P. Hackl, 1978, The Use of MOSUMS for Quality Control, *Technometrics*, **20**: 431-436.
- , and P. Hackl, 1980, An Extension of the MOSUM Technique for Quality Control, *Technometrics*, **22**: 1-7.
- Bowker, A. H., and G. J. Lieberman, 1972, *Engineering Statistics*, 2nd Ed., Prentice Hall, Englewood Cliffs, New Jersey.
- Burr, I. W., 1976, *Statistical Quality Control Methods*, Marcel Dekker, Inc., New York.
- Davies, O. L., 1956, *The Design and Analysis of Industrial Experiments*, Hafner, New York.
- Deming, W. E., 1982, *Quality, Productivity and Competitive Position*, MIT Center for Advanced Engineering Study, Cambridge, Massachusetts.
- Duncan, A. J., 1974, *Quality Control and Industrial Statistics*, 4th Ed., Richard D. Irwin, Illinois.
- Ewan, W. D., 1963, When and How to Use Cu-Sum Charts, *Technometrics*, **5**: 1-22.
- Ferris, C. L., F. R. Grubbs, and C. L. Weaver, 1946, Operating Characteristics for the Common Statistical Tests of Significance, *Annals of Mathematical Statistics*, **17**: 178-197.
- Fisher, R. A., and F. Yates, 1957, *Statistical Tables*, 5th Ed., Oliver and Boyd, Edinburgh.
- , 1960, *The Design of Experiments*, 7th Ed., Oliver and Boyd, Edinburgh.
- Guenther, W. C., 1981, Sample Size Formulas for Normal Theory T Tests, *The American Statistician*, **35**: 243-244.
- Ishikawa, K., 1976, *A Guide to Quality Control*, Asian Productivity Association, Tokyo, Japan. (Available through the American Society for Quality Control, Milwaukee, Wisconsin.)
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- , 1980, Determination of Acceptance Numbers and Sample Sizes for Attribute Sampling Plans, *J. Qual. Technol.*, **12**: 187-190.
- Juran, J. M., and F. M. Gryna, 1980, *Quality Planning and Analysis*, McGraw-Hill, New York.
- Nelson, W., 1982, *Applied Life Data Analysis*, John Wiley & Sons, Inc., New York.
- Rohatgi, V. K., 1984, *Statistical Inference*, John Wiley & Sons, Inc., New York.
- Satterthwaite, F. E., 1946, An Approximate Distribution of Estimates of Variance Components, *Biometrics Bull.*, **2**: 110-114.
- Stephens, L. J., 1978, A Closed Form Solution for Single Sample Acceptance Sampling Plans, *J. Qual. Technol.*, **10**: 159-163.
- Trickett, W. H., B. L. Welch, and G. S. James, 1956, Further Critical Values for the Two Means Problem, *Biometrika*, **43**: 203-205.
- Wadsworth, H. M., K. S. Stevens, and A. B. Godfrey, 1986, *Quality Control*, John Wiley & Sons, Inc., New York.
- Wald, A., 1947, *Sequential Analysis*, John Wiley & Sons, Inc., New York.
- Wetherill, G. B., 1975, *Sequential Methods in Statistics*, 2nd Ed., Chapman and Hall, London.

ADDITIONAL READING

- Bennett, C. A., and N. L. Franklin, 1954, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons, Inc., New York.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter, 1978, *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.
- Guttman, I., S. S. Wilks, and J. S. Hunter, 1982, *Introductory Engineering Statistics*, 3rd Ed., John Wiley & Sons, Inc., New York.
- Hald, A., 1952, *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York.
- Hahn, G. J., and S. S. Shapiro, 1967, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York.

- Hines, W. W., and D. C. Montgomery, 1980, *Probability and Statistics in Engineering and Management Science*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Natrella, M. G., 1960, *Experimental Statistics, National Bureau of Standards Handbook 91*, 2nd Printing, U. S. Government Printing Office, Washington, D. C.

CHAPTER 5

Analysis of Variance

5.0 INTRODUCTION

In nuclear material accounting applications, experiments are often conducted for the purpose of studying and quantifying how certain changes in the environment of an experiment affect particular response variables. This requires that an experiment be conducted with the factors of interest varied in a controlled, prescribed manner, so that the total variability in the resulting response variable data can be partitioned into components corresponding to each factor that was varied in the experiment. It is then possible, through statistical hypothesis testing, to determine which factors cause significant variations in the response variable and whether any of the factors interact to affect the response variable. This information is often used in modeling and characterizing measurement systems and material control and accountability systems. These and other applications are discussed in Chapters 14, 15, and 16.

The statistical methods that are used in planning experiments of this type and in analyzing the resulting data fall under the general headings of **experimental design** and **analysis of variance**. These are very broad subject areas, and it is not possible to thoroughly cover them in this book. Some basic concepts and techniques are presented in this chapter and in Chapter 7 that are adequate for many applications. These concepts and techniques can be extended and adapted to accommodate more complicated applications than are discussed in this book. Careful study of an appropriate text or reference, however, should precede such extensions and adaptations.

This chapter covers estimation and hypothesis testing techniques for analyzing data from some particular types of experiments and observational studies. Chapter 7 covers some elementary methods for planning and executing experiments to ensure that adequate data are collected to meet specific objectives.

In this chapter, **fixed effects** and **random effects** are defined, and methods for analyzing experimental data are presented for each. It should be noted, however, that the random effects methods presented in this chapter assume that the "levels" of a random factor which appear in an experiment are a random sample from an infinite population of levels. In practice, not all populations are infinite. Of those that are finite, some are

large enough to treat as being infinite, while others are not and should be treated as finite. Random effects analysis of variance methods for finite populations are presented by Bennett and Franklin (1954). Rules for converting the estimation procedure for any infinite population situation into one for finite populations are given by Searle and Fawcett (1970).

5.1 ONE-WAY FIXED EFFECTS ANALYSIS OF VARIANCE

Section 4.3.2.1 addresses the problem of comparing the unknown means of two populations when the unknown variances are assumed to be equal. The hypothesis testing procedure in that section assumes that independent random samples are taken from the two populations. The test statistic, given by Equation 4.40, is a function of the difference between the two sample means and a pooled estimate of the common variance. If the null hypothesis of interest is rejected, a confidence interval for the difference in the population means can be constructed by using Equation 3.29 to aid in interpreting the difference. In practice, however, it is often necessary to compare the means of more than two populations. In such situations the **fixed effects analysis of variance** is an appropriate methodology that provides an initial overall test of the equality of two or more means and also provides the quantities necessary for further specific comparisons among the means.

This section introduces the one-way fixed effects analysis of variance with computational formulas for both equal and unequal sample sizes. Power curves, OC curves, and required sample size determination are presented, as well as techniques for making comparisons among the sample means.

5.1.1 An Overall Test for Equal Means

Suppose that it is of interest to compare the means of $a \geq 2$ populations and that independent random samples of size $n > 1$ are taken from each of the populations. The resulting data are symbolically displayed in Table 5.1. Let the subscript i denote population, and let j denote individual sample observations. Then y_{ij} is the j^{th} observed value from population i . Within population i , the observations y_{ij} are assumed to be from a distribution with mean μ_i and variance σ^2 . The means μ_i may differ from population to population, but the variance σ^2 is assumed to be the same for all populations.

In practice, the a samples are sometimes called **classes** or **treatment groups**. These names are descriptive of certain analysis of variance applications. For example, in an experiment involving college students, it might be of interest to compare males and females with respect to some characteristic such as grade point average. Thus, individuals in a random sample of

TABLE 5.1
Sample Data Layout for a
One-Way Analysis of Variance

Population			
1	2	...	a
y_{11}	y_{21}	...	y_{a1}
y_{12}	y_{22}	...	y_{a2}
.	.		
.	.		
y_{1n}	y_{2n}	...	y_{an}

students would be classified by sex, and a comparison of the observed **class means** would provide a basis for reaching a conclusion about the equality of the true population means of the characteristic of interest for male and female college students. The name “treatment groups” comes from applications where a random sample from a homogeneous population is randomly split into a groups, and each group is assigned a different stimulus or treatment. Each treatment group is then assumed to be equivalent to a random sample from the population as if the entire population had been subjected to the corresponding treatment. The objective is to determine the effect of the different treatments on some characteristic of interest. This is accomplished by comparing the **treatment means**. In some applications, the treatments are different levels of a stimulus, such as temperature. Thus, the term **treatment levels** is often used in references to the population or class identifications.

From the above assumptions, the observations can be characterized by the **linear model**

$$y_{ij} = \mu_i + \epsilon_{ij} \quad (5.1)$$

$$i = 1, 2, \dots, a; j = 1, 2, \dots, n$$

where μ_i is the unknown mean of the i^{th} population, and ϵ_{ij} is a **random error** from a distribution with zero mean and with variance σ^2 . It is assumed in the analysis of variance that the ϵ_{ij} are independent observations from a single distribution. The hypothesis testing and interval estimation techniques presented in this section require the additional assumption that the ϵ_{ij} are from a $N(0, \sigma^2)$ distribution. That is, each y_{ij} is assumed to be an observation from a $N(\mu_i, \sigma^2)$ distribution.

Consider now the null hypothesis that all μ_i are equal with the alternative hypothesis that some of the μ_i differ. This is expressed as

$$\begin{aligned} H_0: \mu_1 = \mu_2 = \dots = \mu_a \\ H_A: \text{some of the means are not equal} \end{aligned} \quad (5.2)$$

If we define an overall mean μ , given by

$$\mu = \sum_{i=1}^a \mu_i / a \quad (5.3)$$

then under H_0 every μ_i is equal to μ , and the hypotheses can be written as

$$\begin{aligned} H_0: \mu_i = \mu, \quad i = 1, 2, \dots, a \\ H_A: \mu_i \neq \mu, \quad \text{for some } i \end{aligned} \quad (5.4)$$

This latter representation of H_0 gives rise to another form of the linear model in Equation 5.1. To see this, define the quantity

$$\alpha_i = \mu_i - \mu \quad (5.5)$$

and note that Equation 5.1 can be expressed as

$$\begin{aligned} y_{ij} &= \mu + (\mu_i - \mu) + \epsilon_{ij} \\ &= \mu + \alpha_i + \epsilon_{ij} \\ i &= 1, 2, \dots, a; \quad j = 1, 2, \dots, n \end{aligned} \quad (5.6)$$

In this form of the model, α_i is called the effect of being in the i^{th} population. Thus, each observation y_{ij} is the sum of an **overall mean** μ , a **population effect** α_i , an observation ϵ_{ij} which is assumed to be from a $N(0, \sigma^2)$ distribution.

From Equations 5.3 and 5.5, note that

$$\sum_{i=1}^a \alpha_i = \sum_{i=1}^a (\mu_i - \mu) = \sum_{i=1}^a \mu_i - a\mu = a\mu - a\mu = 0 \quad (5.7)$$

Thus, the hypotheses given by Equations 5.2 and 5.4 can be expressed in terms of the α_i as

$$\begin{aligned} H_0: \alpha_i = 0, \quad i = 1, 2, \dots, a \\ H_A: \alpha_i \neq 0, \quad \text{for some } i \end{aligned} \quad (5.8)$$

That is, if all a populations have the same mean, then each $\mu_i = \mu$ and each $\alpha_i = 0$.

The following notation is introduced to simplify subsequent discussions. Denote the i^{th} sample total by

$$y_i = \sum_{j=1}^n y_{ij} \quad (5.9)$$

Then the i^{th} sample mean is given by

$$\bar{y}_i = y_i/n \quad (5.10)$$

Similarly, denote the overall sample total by

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^n y_{ij} = \sum_{i=1}^a y_i \quad (5.11)$$

Then the overall sample mean is given by

$$\bar{y}_{..} = y_{..}/an \quad (5.12)$$

This so-called "dot notation" is a concise way of representing sums. Replacing a subscript by a dot indicates that the observations are summed over all values of the subscript.

The **total corrected sum of squares** is a measure of the total variability in the data and is given by

$$SS_{\text{Total}} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 \quad (5.13)$$

By analogy with Equation 3.13, SS_{Total} has $(an - 1)$ degrees of freedom. Equation 5.13 can also be expressed as

$$\begin{aligned} SS_{\text{Total}} &= \sum_{i=1}^a \sum_{j=1}^n [(\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i)]^2 \\ &= n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \\ &\quad + 2 \sum_{i=1}^a [(\bar{y}_i - \bar{y}_{..}) \sum_{j=1}^n (y_{ij} - \bar{y}_i)] \end{aligned}$$

But

$$\sum_{j=1}^n (y_{ij} - \bar{y}_i) = n\bar{y}_i - n\bar{y}_i = 0$$

so that the last term in the above expression vanishes, leaving

$$SS_{\text{Total}} = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \quad (5.14)$$

Thus, the total variability in the data, measured by the total corrected sum of squares, can be partitioned and expressed as a sum of two components. The first component, called the **between sample sum of squares**, is a sum of squared differences between the sample means \bar{y}_i and the overall sample mean $\bar{y}_{..}$ and is denoted by

$$SS_{\text{Between}} = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 \quad (5.15)$$

with $(a - 1)$ degrees of freedom. The second component, called the **within sample sum of squares**, is a sum of squared differences between the observations y_{ij} within samples and the sample means \bar{y}_i and is denoted by

$$SS_{\text{Within}} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \quad (5.16)$$

with $a(n - 1)$ degrees of freedom. Thus, Equation 5.14 can be written symbolically as

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}} \quad (5.17)$$

Efficient computational formulas for the sums of squares are obtained by expanding and simplifying the expressions in Equations 5.14, 5.15, and 5.16. The resulting formulas are

$$SS_{\text{Total}} = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an} \quad (5.18)$$

$$SS_{\text{Between}} = \sum_{i=1}^a \frac{y_i^2}{n} - \frac{y_{..}^2}{an} \quad (5.19)$$

and

$$SS_{\text{Within}} = SS_{\text{Total}} - SS_{\text{Between}} \quad (5.20)$$

The between sample mean square and the within sample mean square are obtained by dividing the corresponding sums of squares by their degrees of freedom; that is,

$$MS_{\text{Between}} = SS_{\text{Between}}/(a-1) \quad (5.21)$$

and

$$MS_{\text{Within}} = SS_{\text{Within}}/a(n-1) \quad (5.22)$$

It is from these mean squares that a test of the hypothesis in Equation 5.4 or 5.8 is constructed. It can be shown that the expectations of the mean squares are given by

$$E[MS_{\text{Between}}] = \sigma^2 + n \sum_{i=1}^a (\mu_i - \mu)^2 / (a-1)$$

or, equivalently,

$$E[MS_{\text{Between}}] = \sigma^2 + n \sum_{i=1}^a \alpha_i^2 / (a-1) \quad (5.23)$$

and

$$E[MS_{\text{Within}}] = \sigma^2 \quad (5.24)$$

Thus, MS_{Within} is an unbiased estimator for σ^2 . Note that when $a = 2$, MS_{Within} is the pooled estimator for σ^2 used in the two sample t-tests presented in Section 4.3.2.1 with $n_1 = n_2 = n$. Note also that if the null hypothesis is true, each $(\mu_i - \mu) = 0$; that is, each $\alpha_i = 0$, and Equation 5.23 reduces to $E[MS_{\text{Between}}] = \sigma^2$. This means that under the null hypothesis, MS_{Between} is also an unbiased estimator for σ^2 . When H_0 is

false, however, some of the α_i are not zero, and examination of Equation 5.23 reveals that $E[MS_{\text{Between}}]$ is larger than σ^2 . That is, when H_0 is false, MS_{Between} tends to overestimate σ^2 .

Next, assuming $\epsilon_{ij} \sim N(0, \sigma^2)$, it follows that $y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$, $\bar{y}_{i.} \sim N(\mu + \alpha_i, \sigma^2/n)$, and $\bar{y}_{..} \sim N(\mu, \sigma^2/an)$. Using these results, it can be shown [see Searle (1971) or Graybill (1976) for details] that:

1. $SS_{\text{Within}}/\sigma^2$ has a chi-square distribution with $a(n - 1)$ degrees of freedom.

2. When H_0 is true, $SS_{\text{Between}}/\sigma^2$ has a chi-square distribution with $a - 1$ degrees of freedom.

3. SS_{Between} and SS_{Within} are independent.

Thus, when H_0 is true, the statistic

$$F = \frac{[SS_{\text{Between}}/\sigma^2]/(a-1)}{[SS_{\text{Within}}/\sigma^2]/a(n-1)} \\ = MS_{\text{Between}}/MS_{\text{Within}} \quad (5.25)$$

has an F-distribution with $a - 1$ and $a(n - 1)$ degrees of freedom. When H_0 is false, however, $SS_{\text{Between}}/\sigma^2$ has a **noncentral chi-square distribution** with $a - 1$ degrees of freedom, and F in Equation 5.25 has a **noncentral F-distribution**. The noncentral chi-square and noncentral F-distributions are not discussed in detail in this section. The interested reader is referred to Searle (1971) or Graybill (1976). At the intuitive level, however, the previous discussion implies that if H_0 is false, the numerator of F in Equation 5.25 tends to be larger than σ^2 , while the denominator is an unbiased estimator of σ^2 . Thus, an unusually large value of F suggests that H_0 is false. Specifically, H_0 is rejected at the α level of significance if

$$F \geq F_{1-\alpha}[a-1, a(n-1)] \quad (5.26)$$

where $F_{1-\alpha}[a - 1, a(n - 1)]$ is obtained from Table A6. In practice, the results at each step of the analysis are summarized in an **analysis of variance table** as illustrated in Table 5.2.

Example 5.1 Petit and Kienberger (1961) published the results of a series of experiments to determine whether a particular ignition time and

TABLE 5.2
One-Way Fixed Effects Analysis of Variance Table

Source	df	SS	MS	E[MS]	F
Between samples	$a - 1$	SS_{Between}	MS_{Between}	$\sigma^2 + n \sum_{i=1}^a \alpha_i^2 / (a - 1)$	$\frac{MS_{\text{Between}}}{MS_{\text{Within}}}$
Within samples	$a(n - 1)$	SS_{Within}	MS_{Within}	σ^2	
Total	$an - 1$	SS_{Total}			

temperature combination reliably converts uranium–oxygen compounds to stoichiometric U_3O_8 . Such a conversion procedure is important to the precise determination of uranium content by the gravimetric method and for the preparation of uranium chemical and isotopic standards. One objective of their experimentation was to assess the effect of the starting chemical compound (parent material) on the uranium content of the U_3O_8 resulting from the conversion.

Four 10-g samples of each of three parent materials were ignited for 16 h at 850°C . The data below show the percent uranium content of the resulting U_3O_8 for the 12 samples. Use the one-way analysis of variance to test the hypothesis that the true mean percent uranium is the same for the three parent materials. Use the 0.05 significance level.

Parent Material

	Uranyl fluoride	Uranium peroxide	Uranium trioxide
	84.798	84.794	84.772
	84.793	84.805	84.773
	84.793	84.802	84.760
	84.815	84.797	84.775
$y_{i.}$:	339.199	339.198	339.080
$\bar{y}_{i.}$:	84.7998	84.7995	84.7700
$y_{..} = 1017.477$			

The sums of squares are computed by applying Equations 5.18, 5.19, and 5.20 as follows:

$$\begin{aligned}
 SS_{\text{Total}} &= (84.798^2 + 84.793^2 + \dots + 84.775^2) - \frac{1017.477^2}{12} \\
 &= 86,271.62335 - 86,271.62046 = 0.00289 \\
 SS_{\text{Between}} &= \frac{339.199^2}{4} + \frac{339.198^2}{4} + \frac{339.080^2}{4} - \frac{1017.477^2}{12} \\
 &= 86,271.62280 - 86,271.62046 = 0.00234 \\
 SS_{\text{Within}} &= 0.00289 - 0.00234 = 0.00055
 \end{aligned}$$

The results are summarized in the following analysis of variance table:

Source	df	SS	MS	E[MS]	F
Between materials	2	0.00234	0.00117	$\sigma^2 + \frac{4 \sum_{i=1}^3 \alpha_i^2}{2}$	$\frac{0.00117}{0.0000611} = 19.15$
Within materials	9	0.00055	0.0000611	σ^2	
Total	11	0.00289			

From Table A.6, $F_{0.95}(2,9) = 4.26$; and because $F = 19.15 > 4.26$, reject the hypothesis of equal means.

Methods for further specific comparisons among the means are presented in Section 5.1.3. In this example, however, a preliminary conclusion can be reached by noting that the theoretical percent uranium in stoichiometric U_3O_8 is 84.80, and that the sample means, rounded to three significant digits, are as follows:

Uranyl fluoride:	84.800
Uranium peroxide:	84.800
Uranium trioxide:	84.770

Thus, the sample means for uranyl fluoride and uranium peroxide are identical to the theoretical percent uranium for stoichiometric U_3O_8 , whereas the sample mean for uranium trioxide is somewhat smaller (this obviously caused the rejection of the null hypothesis). A logical conclusion is that igniting for 16 h at 850°C probably does not completely convert uranium trioxide to stoichiometric U_3O_8 .

5.1.2 Unequal Sample Sizes

Often in practice, the number of observations is not the same for all classes. In such cases, the one-way analysis of variance formulas are

slightly different. Let n_i denote the sample size for the i^{th} class; then

$$N = \sum_{i=1}^a n_i \quad (5.27)$$

is the total number of observations. The formulas for the sums of squares and degrees of freedom are given by

$$SS_{\text{Total}} = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \quad (5.28)$$

$$df_{\text{Total}} = N - 1 \quad (5.29)$$

$$SS_{\text{Between}} = \sum_{i=1}^a \frac{y_i^2}{n_i} - \frac{y_{..}^2}{N} \quad (5.30)$$

$$df_{\text{Between}} = a - 1 \quad (5.31)$$

$$SS_{\text{Within}} = SS_{\text{Total}} - SS_{\text{Between}} \quad (5.32)$$

and

$$df_{\text{Within}} = df_{\text{Total}} - df_{\text{Between}} = (N - 1) - (a - 1) = N - a \quad (5.33)$$

As before, a mean square is a sum of squares divided by its degrees of freedom. When the n_i are unequal, the expected value of MS_{Between} is given by

$$E[MS_{\text{Between}}] = \sigma^2 + \sum_{i=1}^a \frac{n_i(\mu_i - \bar{\mu})^2}{a - 1} \quad (5.34)$$

where

$$\bar{\mu} = \sum_{i=1}^a n_i \mu_i / N \quad (5.35)$$

Note that if the μ_i are not equal, $\bar{\mu}$ is not in general equal to the parameter μ , given by Equation 5.3. If H_0 is true, however, each $\mu_i = \mu$, from which it follows that $\bar{\mu} = \mu$, and each $(\mu_i - \bar{\mu})^2 = (\mu - \mu)^2 = 0$. Thus, when H_0 is true, Equation 5.34 reduces to $E[MS_{\text{Between}}] = \sigma^2$. Also, whether or not the sample sizes are equal, $E[MS_{\text{Within}}] = \sigma^2$; that is, MS_{Within} is an unbiased estimator of σ^2 .

As with equal sample sizes, $SS_{\text{Between}}/\sigma^2$ and $SS_{\text{Within}}/\sigma^2$ are independent and have chi-square distributions with $(a - 1)$ and $(N - a)$ degrees of freedom, respectively. Thus, the null hypothesis in Equation 5.4 or 5.8 is rejected at the α level of significance if

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} \geq F_{1-\alpha}(a-1, N-a) \quad (5.36)$$

In constructing an analysis of variance table and performing the hypothesis test of equal means, the major difference caused by unequal sample sizes is in the calculation of sums of squares and degrees of freedom. There are, however, important advantages in having equal sample sizes. First, the test statistic is relatively insensitive to slight departures from the assumption of equal variances if the sample sizes are equal. This may not be true for unequal sample sizes. Second, the power of the F-test is maximized for a fixed total sample size N when the sample sizes are equal. Thus, when possible, the practitioner should design and control an experiment such that the sample sizes are equal, or as nearly so as possible. Experimental design is the topic of Chapter 7, and the reader is referred there for more details.

5.1.3 Power and Required Sample Size

This section describes how power curves and OC curves are constructed and how required sample sizes are determined for the one-way fixed effects analysis of variance with equal sample sizes.

Section 4.1.2 defines Type I and Type II errors and their probabilities of occurring in hypothesis testing. The probability β of a Type II error and the power of a test, $1 - \beta$, are given by Equations 4.3 and 4.4. For the one-way fixed effects analysis of variance, the probability of a Type II error is given by

$$\beta = \Pr[F < F_{1-\alpha}(a-1, N-a) | H_0 \text{ is false}] \quad (5.37)$$

and the power is given by

$$1 - \beta = \Pr[F \geq F_{1-\alpha}(a-1, N-a) | H_0 \text{ is false}] \quad (5.38)$$

Evaluation of Equations 5.37 and 5.38 requires knowing the distribution of F when H_0 is false and specifying the extent to which the means μ_i differ. From the discussion in Section 5.1.1, when H_0 is false, F has a noncentral F -distribution with $a - 1$ and $N - a$ degrees of freedom and with noncentrality parameter λ , given by

$$\lambda = \left(n \sum_{i=1}^a (\mu_i - \mu)^2 / \sigma^2 \right)^{1/2} \quad (5.39)$$

Note that if H_0 is true, $\lambda = 0$. As λ increases, however, β in Equation 5.37 decreases, and $1 - \beta$ in Equation 5.38 increases.

Probabilities are difficult to compute directly from the noncentral F -distribution. However, Pearson and Hartley (1951) provide charts that give the power, $1 - \beta$, of an F -test for various combinations of α , $\nu_1 = (a - 1)$, $\nu_2 = (N - a)$, and ϕ , where ϕ is given by

$$\phi = \lambda / \sqrt{a} = \left(n \sum_{i=1}^a (\mu_i - \mu)^2 / a \sigma^2 \right)^{1/2} \quad (5.40)$$

A subset of these charts is reproduced in Table A7 for values of $\nu_1 = 1$ through $\nu_1 = 8$, with $\alpha = 0.05$ and $\alpha = 0.01$.

Examination of ϕ in Equation 5.40 reveals that the use of the Pearson-Hartley charts, and the interpretation of the resulting probabilities, depends upon the interpretation of the quantity

$$\sum_{i=1}^a (\mu_i - \mu)^2$$

Suppose that it is of interest to determine the power of a test when at least two of the means differ by at least a specified amount δ . Rather than determining the power for every possibility (i.e., two means differ, three means differ, and so on), a conservative approach is to consider the case where two of the means differ by an amount δ , and the remaining $(a - 2)$ means are equal to the average of the two that differ. Then $(\mu_i - \mu)^2$ is equal to $(\delta/2)^2$ for the two extreme μ_i , and equal to zero for the others. In this case,

$$\sum_{i=1}^a (\mu_i - \mu)^2 = 2(\delta/2)^2 = \delta^2/2$$

which is substituted into Equation 5.40 to obtain

$$\phi = \sqrt{n \delta^2 / 2a \sigma^2} \quad (5.41)$$

Usually in determining power, σ^2 is unknown. If the purpose is to construct a power curve or an OC curve for an analysis of variance that is already completed, then MS_{within} is substituted for σ^2 in Equation 5.40 or 5.41. However, if the purpose is to provide guidance in choosing a sample size n for an experiment that is being planned, then σ^2 must be estimated from similar previous studies or guesswork based on scientific laws. Unfortunately, a poor choice of σ^2 can yield misleading results. This is illustrated by Scheffé (1959). An often satisfactory alternative is to express δ as a multiple of σ . For example, it might be of interest to detect when two of the means differ by $\delta = 0.5\sigma$. Then substituting 0.5σ for δ in Equation 5.41 gives

$$\phi = \sqrt{n(0.5)^2 / 2a} = \sqrt{n/8a}$$

which does not require an estimate of σ^2 .

Use of the charts in Table A7 is best illustrated by example. Thus, the following two examples cover the construction of a power curve and an OC curve for an analysis of variance that has been completed, and the determination of the required sample size n for an experiment that is being planned.

Example 5.2 Suppose that in Example 5.1 we adopt the conservative philosophy leading to Equation 5.41. Determine the points for constructing a power curve and an OC curve for the F-test in Example 5.1.

First, note that in Example 5.1, $\alpha = 0.05$, $a = 3$, $n = 4$, and $MS_{\text{within}} = 0.0000611$. Thus, the following quantities are needed to use Table A7:

$$\alpha = 0.05$$

$$\nu_1 = a - 1 = 2$$

$$\nu_2 = na - a = 12 - 3 = 9$$

and

$$\phi \cong [4\delta^2/2(3)(0.0000611)]^{1/2} = (104.456)\delta$$

Entering Table A7 with $\alpha = 0.05$, $\nu_1 = 2$, and $\nu_2 = 9$, the corresponding curve is a plot of $1 - \beta$ against ϕ . Because the objective is to plot either $1 - \beta$ or β against δ , note that $\delta = \phi/104.456$ and construct the following table.

ϕ	δ	$1 - \beta$	β
0.0	0.0	0.05	0.95
1.0	0.0096	0.23	0.77
1.2	0.0115	0.32	0.68
1.4	0.0134	0.43	0.57
1.6	0.0153	0.55	0.45
1.8	0.0172	0.65	0.35
2.0	0.0191	0.75	0.25
2.2	0.0211	0.82	0.18
2.4	0.0230	0.89	0.11
2.6	0.0249	0.93	0.07
2.8	0.0268	0.96	0.04
3.0	0.0287	0.978	0.022
3.2	0.0306	0.988	0.012

The resulting power curve and OC curve are constructed by plotting $1 - \beta$ against δ and then β against δ , respectively. If an acceptable power is $1 - \beta \geq 0.82$, then the experiment described in Example 5.1 and the resulting analysis of variance are satisfactory for detecting a difference of approximately 0.0211 or larger between any two of the true means. Conceptually, this means that if the experiment described in Example 5.1 could be repeated a large number of times and if any two of the true means differ by about 0.0211 or more, then the F-test would be expected to reject the hypothesis of equal class means at least 82% of the time.

Example 5.3 Suppose that the experiment described in Example 5.1 is only a pilot study and that the full-scale experiment will compare seven parent materials. It is of interest to detect differences among the true means of size $\delta = 0.015$ with probability 0.90 when testing at the $\alpha = 0.05$ level of significance. Determine the sample size n that satisfies these requirements.

The necessary quantities for using Table A7 are as follows:

$$1 - \beta = 0.90$$

$$\alpha = 0.05$$

$$\nu_1 = a - 1 = 7 - 1 = 6$$

$$\nu_2 = na - a = a(n - 1) = 7(n - 1)$$

$$\sigma^2 \cong MS_{\text{within}} = 0.0000611$$

and, from Equation 5.41,

$$\phi \cong \sqrt{n(0.015)^2/2(7)(0.0000611)} = 0.5129\sqrt{n}$$

or

$$n \cong (\phi/0.5129)^2$$

Because the objective is to determine n , enter Table A7 with $\nu_1 = 6$ and $\alpha = 0.05$. To get a first approximation of n , look at the curve for $\nu_2 = \infty$ and note that $1 - \beta = 0.90$ when $\phi \cong 1.6$. Substituting this value of ϕ into the above expression for n gives $n \cong (1.6/0.5129)^2 = 9.73 \cong 10$. A first approximation of n obtained from the curve for $\nu_2 = \infty$ is always too small because for any finite n the actual curve is to the right of the curve for $\nu_2 = \infty$, thus requiring a larger value of n for a specified power.

If $n = 10$, then $\nu_2 = 7(9) = 63$. Because in Table A7 there are no curves for values of ν_2 between 60 and ∞ , it is not possible to use interpolation to find a value of ϕ for $1 - \beta = 0.90$ at $\nu_2 = 63$. For any value of $1 - \beta$, however, the curve for $\nu_2 = 60$ yields a larger value of ϕ , and thus a larger required sample size, than would be obtained for any $\nu_2 > 60$. Thus, using the curve for $\nu_2 = 60$ provides a conservative estimate of the required sample size for this example.

The curve for $\nu_2 = 60$ indicates that $\phi \cong 1.68$ when $1 - \beta = 0.90$. Substituting this value of ϕ into the above equation yields $n \cong (1.68/0.5129)^2 = 10.73 \cong 11$. Thus, the experiment should include $n = 11$ samples of each of the seven parent materials, for a total of $na = 77$ material samples.

5.1.4 Estimation and Comparison of Class Means

Constructing an analysis of variance table and performing an overall test of equal means, as described in Sections 5.1.1 and 5.1.2, is only the

initial phase of a complete analysis of the data. The next phase involves examining and making comparisons among the class sample means. This section deals with constructing confidence intervals for the unknown true class means and three basic types of comparisons: planned comparisons, comparisons among all possible pairs of class sample means, and comparisons suggested by the data. Each of the three types of comparisons has its own anomalies and requires its own specialized statistical procedures.

5.1.4.1 Confidence Intervals for the Class Means

A summary display of results is an essential part of an effective data analysis. In addition to an analysis of variance table, such as Table 5.2, a list of the class sample means and their estimated standard deviations is essential.

In the one-way fixed effects analysis of variance, it is assumed that y_{ij} is from a $N(\mu_i, \sigma^2)$ distribution. Thus, the i^{th} sample mean \bar{y}_i is from a $N(\mu_i, \sigma^2/n_i)$ distribution. Because σ^2 is assumed to be the same for all classes, a pooled estimator is given by

$$S^2 = MS_{\text{within}} \quad (5.42)$$

which has $N - a$ degrees of freedom.

A $100(1 - \alpha)\%$ confidence interval for μ_i is given by

$$\bar{y}_i \pm t_{1-\alpha/2}(N-a) \sqrt{S^2/n_i} \quad (5.43)$$

where $t_{1-\alpha/2}(N - a)$ is obtained from Table A5.

Example 5.4 Using the results from Example 5.1, construct 95% confidence intervals for the three unknown true means.

The following quantities are needed for application of Equation 5.43:

$$\sqrt{S^2/n_i} = \sqrt{MS_{\text{within}}/n_i} = \sqrt{0.0000611/4} = 0.003908$$

$$t_{1-\alpha/2}(N-a) = t_{0.975}(9) = 2.26$$

$$\bar{y}_1 = 84.7998$$

$$\bar{y}_2 = 84.7995$$

$$\bar{y}_3 = 84.7700$$

By substituting these quantities into Equation 5.43, the 95% confidence intervals are

$$\begin{aligned}\mu_1: & 84.7998 \pm 2.26(0.003908) \\ & = 84.7998 \pm 0.0088 \\ & \text{or } (84.7910, 84.8086) \\ \mu_2: & (84.7907, 84.8083) \\ \mu_3: & (84.7612, 84.7788)\end{aligned}$$

Note that the value 84.80, which corresponds to stoichiometric U_3O_8 , is inside the confidence intervals for μ_1 and μ_2 but outside the confidence interval for μ_3 . This should interest the investigator because it implies that μ_3 is less than 84.80 or that the ignition conditions used in the experiment will not completely convert uranium trioxide to stoichiometric U_3O_8 .

Any inferences based upon confidence intervals for individual means should be made and interpreted with caution. To see why, recall that the data are assumed to be independent random samples from a normal distributions. If $100(1 - \alpha)\%$ confidence intervals are constructed for all a of the μ_i , then the probability that all a intervals simultaneously cover their respective μ_i is approximately $(1 - \alpha)^a$ (Definition 2.12). In Example 5.4, the probability that all three 95% confidence intervals simultaneously cover their respective true means is approximately $(0.95)^3 = 0.8574$. Thus, the approximate probability that at least one of the intervals does not cover its true mean is $1 - (1 - \alpha)^a$, which in Example 5.4 is $1 - (0.95)^3 = 0.1426$. As the number of intervals increases, the overall probability of reaching a wrong conclusion also increases. For example, if individual 95% confidence intervals are constructed for $a = 10$ means, then $(0.95)^{10} = 0.5987$ and $1 - (0.95)^{10} = 0.4013$. That is, the approximate probability is 0.4013 that at least one of the intervals does not cover its true mean.

There are a number of possible remedies for this situation. Two are considered here. The first is a special case of a procedure introduced by Scheffé (1959). Rather than using Equation 5.43, the formula

$$\bar{y}_i \pm \sqrt{(a-1)F_{1-\alpha}(a-1, N-a)} \sqrt{S^2/n_i} \quad (5.44)$$

yields confidence intervals for the a unknown true class means with an overall probability of at least $(1 - \alpha)$ that the a intervals simultaneously

cover their respective true means. Note that $F_{1-\alpha}(a-1, N-a)$ is obtained from Table A6.

The second procedure uses an inequality due to Bonferroni [no reference available, see Miller (1966) for a discussion] which basically guarantees that using α/a in place of α in Equation 5.43 will yield confidence intervals for the a unknown true class means with an overall probability of at least $(1 - \alpha)$ that the a intervals simultaneously cover their respective true means. The appropriate formula is thus

$$\bar{y}_i \pm t_{1-\alpha/2a}(N-a) \sqrt{S^2/n_i} \quad (5.45)$$

Either method guarantees an overall confidence level of at least $(1 - \alpha)$. The width of the resulting intervals, however, will not in general be the same for both methods. Thus, it may be desirable to construct intervals by both methods and then report the narrower intervals. That is, because both methods guarantee the same overall confidence level, it is overly conservative to use intervals that are wider than necessary.

Example 5.5 Using the results from Example 5.1, construct confidence intervals for the three unknown true means, with an overall confidence level of 95%. Use both the Scheffé method and the Bonferroni method and compare the results.

By using Equation 5.44 with $F_{0.95}(2,9) = 4.26$, the Scheffé confidence intervals are given by

$$\bar{y}_i \pm \sqrt{2(4.26)} \sqrt{0.0000611/4}$$

or

$$\bar{y}_i \pm 0.0114$$

Thus, we can be 95% confident that the following intervals simultaneously cover the respective true class means:

$$\mu_1: (84.7884, 84.8112)$$

$$\mu_2: (84.7881, 84.8109)$$

$$\mu_3: (84.7586, 84.7814)$$

By using Equation 5.45 with $t_{0.9917}(9) \cong 2.966$ (computed by linear interpolation), the Bonferroni confidence intervals are given by

$$\bar{y}_i \pm (2.966) \sqrt{0.0000611/4}$$

or

$$\bar{y}_i \pm 0.0116$$

Thus, we can be 95% confident that the following intervals simultaneously cover the respective true class means:

$$\mu_1 : (84.7882, 84.8114)$$

$$\mu_2 : (84.7879, 84.8111)$$

$$\mu_3 : (84.7584, 84.7816)$$

Note that the Bonferroni intervals are slightly wider than the Scheffé intervals. In this example, however, the discrepancy is of no practical importance. Because neither of the intervals for μ_3 includes 84.80, we conclude that the ignition conditions used in the experiment probably will not completely convert uranium trioxide to stoichiometric U_3O_8 .

5.1.4.2 Planned Comparisons

In controlled experiments, the investigator typically plans an experiment in such a way that certain comparisons among the class means can be made. For instance, suppose that in Example 5.1 the investigator is reasonably certain that igniting uranyl fluoride (UF) for 16 h at 850°C will, on the average, convert this material to stoichiometric U_3O_8 . The investigator is not sure, however, what will happen when either uranium peroxide (UP) or uranium trioxide (UT) is ignited under the same conditions. Thus, the investigator quite likely had two specific comparisons in mind when he planned the experiment:

1. UP vs UF

2. UT vs UF

If μ_1 , μ_2 , and μ_3 denote the true mean percent uranium content of U_3O_8 converted from UF, UP, and UT, respectively, then the above comparisons can be stated as the following statistical hypotheses:

$$1. H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \neq 0$$

$$\begin{aligned} 2. \quad H_0 : \mu_1 - \mu_3 &= 0 \\ H_A : \mu_1 - \mu_3 &\neq 0 \end{aligned}$$

Note that the null hypothesis in each case above is expressed as a linear combination of the true means, which can be written symbolically as

$$C = \sum_{i=1}^a C_i \mu_i \quad (5.46)$$

where the C_i are specified constants. That is, the first null hypothesis has $C_1 = 1$, $C_2 = -1$, and $C_3 = 0$; whereas the second null hypothesis has $C_1 = 1$, $C_2 = 0$, and $C_3 = -1$. Note also that in each case

$$\sum_{i=1}^a C_i = 0 \quad (5.47)$$

In general, any linear combination of means having the form given by Equation 5.46, where the C_i satisfy Equation 5.47, is called a **contrast** of the means. Note that if the null hypothesis of equal class means given by Equation 5.4 is true, then each $\mu_i = \mu$, and Equation 5.46 reduces to

$$C = \sum_{i=1}^a C_i \mu_i = \mu \sum_{i=1}^a C_i$$

which is $\mu(0) = 0$ if Equation 5.47 holds.

Also note that to work in terms of α_i rather than μ_i , Equation 5.46 can be expressed as

$$\begin{aligned} C &= \sum_{i=1}^a C_i \mu_i = \sum_{i=1}^a C_i (\mu + \alpha_i) \\ &= \mu \left(\sum_{i=1}^a C_i \right) + \sum_{i=1}^a C_i \alpha_i \\ &= \sum_{i=1}^a C_i \alpha_i \end{aligned} \quad (5.48)$$

Whether working in terms of Equation 5.46 or 5.48, an estimator of C is obtained by substituting \bar{y}_i for μ_i , which gives

$$\hat{C} = \sum_{i=1}^a C_i \bar{y}_i \quad (5.49)$$

Because each \bar{y}_i is from a $N(\mu_i, \sigma^2/n_i)$ distribution, it follows (see Section 2.7.1) that

$$\hat{C} \sim N \left(\sum_{i=1}^a C_i \mu_i, \sigma^2 \sum_{i=1}^a C_i^2 / n_i \right) \quad (5.50)$$

An estimator of the variance of \hat{C} is obtained by substituting MS_{Within} for σ^2 in Equation 5.50. This estimator is denoted by $V\hat{a}r(\hat{C})$ and is given by

$$V\hat{a}r(\hat{C}) = MS_{\text{Within}} \sum_{i=1}^a C_i^2 / n_i \quad (5.51)$$

Consider testing the hypothesis

$$H_0: \sum_{i=1}^a C_i \mu_i = 0$$

against the alternative

(5.52)

$$H_A: \sum_{i=1}^a C_i \mu_i = C_A \neq 0$$

at the α level of significance. If H_0 is true, the statistic

$$t = \frac{\hat{C}}{\sqrt{V\hat{a}r(\hat{C})}} \quad (5.53)$$

has a Student's t-distribution with $N-a$ degrees of freedom. Thus, H_0 is rejected if

$$t \geq t_{1-\alpha/2}(N-a)$$

or

(5.54)

$$t \leq -t_{1-\alpha/2}(N-a)$$

where $t_{1-\alpha/2}(N-a)$ is obtained from Table A5.

A $100(1-\alpha)\%$ confidence interval for C is given by

$$\hat{C} \pm \sqrt{\hat{V}\hat{\text{ar}}(\hat{C})} t_{1-\alpha/2}(N-a) \quad (5.55)$$

Example 5.6 Suppose that prior to the experiment described in Example 5.1, the investigator was reasonably certain (from the results of previous experiments) that igniting uranyl fluoride (UF) for 16 h at 850°C would, on the average, convert this material to stoichiometric U_3O_8 . The investigator was not sure, however, what would happen if either uranium peroxide (UP) or uranium trioxide (UT) was ignited under the same conditions. Thus, suppose that the investigator conducted the experiment to compare UP with UF and UT with UF. To be consistent with the data displayed in Example 5.1, let μ_1 , μ_2 , and μ_3 denote the true mean percent uranium for UF, UP, and UT, respectively. By defining the contrasts

$$C_1 = \mu_1 - \mu_2$$

and

$$C_2 = \mu_1 - \mu_3$$

the hypotheses of interest are

1. $H_0: C_1 = 0$
 $H_A: C_1 \neq 0$
2. $H_0: C_2 = 0$
 $H_A: C_2 \neq 0$

These are each to be tested at the $\alpha = 0.05$ level of significance.

By using the results from Example 5.1 with Equation 5.49, estimates of C_1 and C_2 are

$$\hat{C}_1 = \bar{y}_1 - \bar{y}_2 = 84.7998 - 84.7995 = 0.0003$$

and

$$\hat{C}_2 = \bar{y}_1 - \bar{y}_3 = 84.7998 - 84.7700 = 0.0298$$

By applying Equation 5.51, estimates of the variances of \hat{C}_1 and \hat{C}_2 are

$$\begin{aligned}\text{Vâr}(\hat{C}_1) &= \text{Vâr}(\hat{C}_2) = (0.0000611) \left[\frac{1^2}{4} + \frac{(-1)^2}{4} \right] \\ &= 0.00003055\end{aligned}$$

and estimates of the standard deviations are

$$\sqrt{\text{Vâr}(\hat{C}_1)} = \sqrt{\text{Vâr}(\hat{C}_2)} = \sqrt{0.00003055} = 0.005527$$

The test statistics, computed from Equation 5.53, are

$$t_1 = \frac{\hat{C}_1}{\sqrt{\text{Vâr}(\hat{C}_1)}} = 0.0003/0.005527 = 0.054$$

and

$$t_2 = \frac{\hat{C}_2}{\sqrt{\text{Vâr}(\hat{C}_2)}} = 0.0298/0.005527 = 5.39$$

From Table A5, $t_{0.975}(9) = 2.26$. Since $-2.26 < t_1 = 0.054 < 2.26$, there is insufficient evidence to conclude that $C_1 \neq 0$. Because $t_2 = 5.39 > 2.26$, however, the hypothesis that $C_2 = 0$ is rejected, and we conclude that $C_2 > 0$. In terms of the true mean percent uranium values, there is insufficient evidence to conclude that μ_1 and μ_2 are different, but there is strong evidence that μ_3 is smaller than μ_1 . This leads to the conclusion that igniting for 16 h at 850°C does not completely convert UT to stoichiometric U_3O_8 .

5.1.4.3 Duncan's Multiple Range Test

Sometimes in practice an investigator cannot specify in advance which contrasts of the means are to be tested. In such cases, the purpose of the experiment may be to attempt to identify which of the population means differ from one another. This can be accomplished by considering every possible pair of the sample means, where for each pair the hypothesis of equal population means is tested. That is, the hypothesis

$$H_0 : \mu_i = \mu_j$$

is tested against the alternative

$$H_A : \mu_i \neq \mu_j$$

for each of the $a(a-1)/2$ pairs of means where $i \neq j$. If, however, the testing method presented in Section 4.3.2.1 is used to test each pair of means at the α level of significance, then the probability of incorrectly rejecting at least one of the null hypotheses is larger than α when $a > 2$. From the Bonferroni inequality discussed in Section 5.1.4.1, an upper bound on this probability is given by

$$1 - (1 - \alpha)^{[a(a-1)/2]}$$

For example, if $\alpha = 0.05$ and $a = 5$, then there are $5(4)/2 = 10$ pairs of means, and the probability of incorrectly rejecting at least one of the null hypotheses could be as large as

$$1 - (0.95)^{10} = 0.4013$$

There are several valid procedures that avoid this problem. Among the more popular of these are the Newman-Keuls test, see Newman (1939) and Keuls (1952) or Hicks (1982); Tukey's test, see Tukey (1949), Box, Hunter, and Hunter (1978), or Graybill (1976); and Duncan's multiple range test, see Duncan (1955). This section presents Duncan's multiple range test.

To apply Duncan's multiple range test when the sample sizes are equal, the a class sample means are arranged in ascending order. The estimated standard deviation of each sample mean is given by

$$S_{\bar{y}_i} = \sqrt{MS_{\text{within}}/n} \quad (5.56)$$

where n is the sample size for each class. From Table A8, obtain the values $q_\alpha(p, N-a)$, for $p = 2, \dots, a$, where α is the desired level of significance and $N-a$ is the degrees of freedom for MS_{within} . For each p , the **least significant range** R_p is calculated as

$$R_p = q_\alpha(p, N-a) \sqrt{MS_{\text{within}}/n}, \quad p = 2, 3, \dots, a \quad (5.57)$$

Then the ranges (or differences) between sample means are tested, beginning with the largest sample mean minus the smallest, which is compared with R_a . If the range is larger than R_a , the two sample means are declared significantly different, which implies that the corresponding true means are different. Next, the range given by the largest sample mean minus the second smallest is compared with R_{a-1} . This continues until all sample means have been compared with the largest. The range given by the second largest sample mean minus the smallest is then compared with R_{a-1} . The second largest sample mean minus the second smallest is compared with R_{a-2} , and so on. This process continues until all $a(a-1)/2$ possible pairs of sample means have been tested.

It sometimes happens that two sample means that are declared significantly different fall between two sample means that are not significantly different. To avoid contradictions, a difference between two means is considered nonsignificant if the two sample means fall between two other means that do not differ significantly.

When the sample sizes are unequal, an approximate procedure can be used if the n_i are not too different. In the approximate procedure, the harmonic mean \bar{n} of the sample sizes, given by

$$\bar{n} = a / (1/n_1 + 1/n_2 + \dots + 1/n_a) \quad (5.58)$$

is substituted for n in Equation 5.56. That is, a pseudo standard deviation is computed as

$$S_{\bar{y}_i}^* = \sqrt{MS_{\text{within}} / \bar{n}} \quad (5.59)$$

Then Equation 5.57 for computing the least significant ranges is appropriately modified, to the form

$$R_p^* = q_\alpha(p, N-a) S_{\bar{y}_i}^* \quad (5.60)$$

The testing then proceeds exactly like Duncan's multiple range test, except that the R_p^* values are used in place of the R_p values. This method is briefly mentioned by Snedecor and Cochran (1980), but they give no specific reference.

Example 5.7 The experiment described in Example 5.1 actually involved more parent materials and ignition conditions than the three used

in that example. The data for five classes (i.e., five sets of test conditions) are given in the following table:

Experimental Data					
Class	1	2	3	4	5
Parent material	Metallic uranium	Uranyl fluoride	Uranyl fluoride	Uranium peroxide	Uranium trioxide
Ignition conditions	2 h at 850°C	16 h at 850°C	4 h at 900°C	16 h at 850°C	16 h at 850°C
	84.794	84.798	84.795	84.794	84.772
	84.792	84.793	84.798	84.805	84.773
	84.790	84.793	84.802	84.802	84.760
		84.815	84.805	84.797	84.775
y_i	254.376	339.199	339.200	339.198	339.080
\bar{y}_i	84.7920	84.7998	84.8000	84.7995	84.7700
n_i	3	4	4	4	4

The total sample size, computed from Equation 5.27, is

$$N = \sum_{i=1}^5 n_i = 19$$

Application of the formulas in Section 5.1.2 for unequal sample sizes yields the one-way analysis of variance results summarized below.

Source	df	SS	MS	F
Between classes	4	0.00265593	0.000663983	15.40
Within classes	14	0.00060375	0.000043125	
Total	18	0.00325968		

Table A6 does not give a value of $F_{0.95}(4,14)$; however, the values $F_{0.95}(4,12) = 3.26$ and $F_{0.95}(4,15) = 3.06$ are given and can be used with linear interpolation to compute

$$F_{0.95}(4,14) \cong 3.06 + (3.26 - 3.06) \frac{(15-14)}{(15-12)} = 3.13$$

Because $F = 15.40 > 3.13$, the hypothesis of equal means is rejected.

Next, apply the approximate multiple range procedure, with $\alpha = 0.05$, to compare all possible pairs of the class sample means.

The harmonic mean of the sample sizes, computed from Equation 5.58, is

$$\bar{n} = 5 / (1/3 + 1/4 + 1/4 + 1/4 + 1/4) = 3.75$$

Substituting \bar{n} and MS_{Within} into Equation 5.59 gives

$$S_{\bar{y}_i}^* = \sqrt{0.000043125/3.75} = 0.00339$$

Entering Table A8 with $\alpha = 0.05$ and $N - a = 14$, the values

$$q_{0.05}(2, 14) = 3.03$$

$$q_{0.05}(3, 14) = 3.18$$

$$q_{0.05}(4, 14) = 3.27$$

$$q_{0.05}(5, 14) = 3.33$$

are obtained, and Equation 5.60 is used to compute the values

$$R_2^* = 3.03 (0.00339) = 0.01027$$

$$R_3^* = 3.18 (0.00339) = 0.01078$$

$$R_4^* = 3.27 (0.00339) = 0.01109$$

$$R_5^* = 3.33 (0.00339) = 0.01129$$

The sample means are arranged in ascending order as shown below:

$$\bar{y}_5 = 84.7700$$

$$\bar{y}_1 = 84.7920$$

$$\bar{y}_4 = 84.7995$$

$$\bar{y}_2 = 84.7998$$

$$\bar{y}_3 = 84.8000$$

The testing then proceeds as follows:

$$\bar{y}_3 - \bar{y}_5 = 0.0300 > R_5^*$$

$$\bar{y}_3 - \bar{y}_1 = 0.0080 < R_4^*$$

$$\bar{y}_3 - \bar{y}_4 = 0.0005 < R_3^*$$

$$\bar{y}_3 - \bar{y}_2 = 0.0002 < R_2^*$$

$$\bar{y}_2 - \bar{y}_5 = 0.0298 > R_4^*$$

$$\bar{y}_2 - \bar{y}_1 = 0.0078 < R_3^*$$

$$\bar{y}_2 - \bar{y}_4 = 0.0003 < R_2^*$$

$$\bar{y}_4 - \bar{y}_5 = 0.0295 > R_3^*$$

$$\bar{y}_4 - \bar{y}_1 = 0.0075 < R_2^*$$

$$\bar{y}_1 - \bar{y}_5 = 0.0220 > R_2^*$$

These results indicate that \bar{y}_5 is significantly less than each of the other four sample means. There are no significant differences, however, among the sample means for classes 1, 2, 3, and 4. This implies that μ_5 is less than the other four true mean percent uranium values, but there is not sufficient evidence to conclude that μ_1 , μ_2 , μ_3 , and μ_4 differ from one another.

It should be noted that short cuts can be utilized with Duncan's multiple range test to minimize the total number of calculations. In this example, the first two comparisons indicate that $\bar{y}_3 - \bar{y}_5$ is significant, but $\bar{y}_3 - \bar{y}_1$ is not. Because \bar{y}_4 and \bar{y}_2 are between \bar{y}_1 and \bar{y}_3 , there can be no significant differences among \bar{y}_1 , \bar{y}_4 , \bar{y}_2 , and \bar{y}_3 ; thus, further tests among these four sample means are not necessary. All that remains is to determine whether \bar{y}_5 is significantly smaller than \bar{y}_1 , \bar{y}_4 , and \bar{y}_2 . Because \bar{y}_1 is closest to \bar{y}_5 , the testing could begin with $\bar{y}_1 - \bar{y}_5$, which is significant. Because both \bar{y}_4 and \bar{y}_2 are greater than \bar{y}_1 , they are certainly significantly greater than \bar{y}_5 , so that $\bar{y}_4 - \bar{y}_5$ and $\bar{y}_2 - \bar{y}_5$ need not be tested. Using this approach, the above example requires testing only 3 of the 10 possible differences.

5.1.4.4 Comparisons Suggested by the Data

The testing procedure presented in Section 5.1.4.2 is recommended for any comparisons among the class sample means that the experiment was designed to make. Sometimes, however, in examining the means, whether by an "eyeball inspection" or by a procedure such as Duncan's multiple range test, patterns are noticed which suggest certain comparisons that were not planned in advance. The corresponding contrasts are estimated by using Equation 5.49, but the t-test given by Equation 5.54 is not valid because the contrasts were selected after examining the data. Thus, when using the t-test, the probability of finding an erroneous significant result (i.e., making a Type 1 error) is larger than α , the specified level of significance. Scheffé (1959) provides a conservative test for this situation. By using Scheffé's test, any number of comparisons can be simultaneously tested, and the overall probability of making a Type I error is at most α .

Scheffé's test requires computing a test statistic t from Equation 5.53 for each contrast. The null hypothesis H_0 , given by Equation 5.52, is rejected for a particular contrast if either

$$t \geq \sqrt{(a-1)F_{1-\alpha}(a-1, N-a)}$$

or

(5.61)

$$t \leq -\sqrt{(a-1)F_{1-\alpha}(a-1, N-a)}$$

where a is the number of classes in the experiment, $N-a$ is the degrees of freedom for MS_{within} , and $F_{1-\alpha}(a-1, N-a)$ is obtained from Table A6.

Confidence intervals for the unknown true values of any number of contrasts can be constructed from the formula

$$\hat{C} \pm \sqrt{\hat{V}\hat{a}r(\hat{C})} \sqrt{(a-1)F_{1-\alpha}(a-1, N-a)} \quad (5.62)$$

where \hat{C} and $\hat{V}\hat{a}r(\hat{C})$ are given by Equations 5.49 and 5.51. The overall probability that these intervals simultaneously cover their true parameter values is at least $1-\alpha$.

When $a = 2$, Scheffé's test and confidence intervals are identical to the t -test and confidence intervals that result from Equations 5.54 and 5.55. This is because

$$\sqrt{F_{1-\alpha}(1, N-a)} = t_{1-\alpha/2}(N-a)$$

When $a > 2$, however, a substantially larger (positive or negative) value of t is required to reject H_0 with Scheffé's test than with the t -test. For example, let $\alpha = 0.05$, and suppose that there are a total of $N = 32$ observations in an experiment. If there are only $a = 2$ classes in the experiment, then $t_{0.975}(30) = 2.04$ and

$$\sqrt{(2-1)F_{0.95}(1, 30)} = \sqrt{4.17} = 2.04$$

However, if there are $a = 8$ classes in the experiment, then $t_{0.975}(24) = 2.06$ and

$$\sqrt{(7)F_{0.95}(7, 24)} = \sqrt{(7)2.42} = 4.12$$

Thus, when $a > 2$, Scheffé's test gives protection against Type I errors, but at the expense of less powerful tests.

Example 5.8 Suppose that in Example 5.7 the experimenter wishes to test the hypothesis that all of the following contrasts are zero with an overall significance level of $\alpha = 0.10$:

$$C_1 : (\mu_2 + \mu_3)/2 - \mu_1$$

$$C_2 : (\mu_2 + \mu_3)/2 - \mu_4$$

$$C_3 : (\mu_2 + \mu_3)/2 - \mu_5$$

$$C_4 : (\mu_2 + \mu_3 + \mu_4)/3 - \mu_1$$

$$C_5 : (\mu_2 + \mu_3 + \mu_4)/3 - \mu_5$$

$$C_6 : (\mu_1 + \mu_2 + \mu_3 + \mu_4)/4 - \mu_5$$

From Example 5.7, $a = 5$ and $N - a = 14$. Table A6 does not give a value of $F_{0.90}(4,14)$; however, the values $F_{0.90}(4,12) = 2.48$ and $F_{0.90}(4, 15) = 2.36$ are given and can be used with linear interpolation to compute

$$F_{0.90}(4,14) \cong 2.36 + (2.48 - 2.36) \frac{(15-14)}{(15-12)} = 2.40$$

Thus,

$$\sqrt{(5-1)F_{0.90}(4,14)} = \sqrt{(4)2.40} = 3.10$$

Estimates and test statistics obtained by applying Equations 5.49, 5.51, and 5.53 are displayed below. The reader is encouraged to verify some of these values.

Contrast	\hat{C}	$\sqrt{\hat{\text{Var}}(\hat{C})}$	t	
C_1	0.0079	0.00445	1.78	< 3.10
C_2	0.0004	0.00402	0.10	< 3.10
C_3	0.0299	0.00402	7.44	> 3.10
C_4	0.0078	0.00424	1.84	< 3.10
C_5	0.0298	0.00379	7.86	> 3.10
C_6	0.0278	0.00370	7.51	> 3.10

The t statistics for the estimated contrasts \hat{C}_3 , \hat{C}_5 , and \hat{C}_6 exceed 3.10; thus, these estimated contrasts are significantly larger than zero, which implies that the unknown true contrasts are larger than zero. The estimated contrasts \hat{C}_1 , \hat{C}_2 , and \hat{C}_4 are not significantly different than zero.

The estimated contrasts \hat{C}_1 , \hat{C}_2 , and \hat{C}_3 compare the average of \bar{y}_2 and \bar{y}_3 against each of the other three sample means, and only \bar{y}_5 is significantly different (less). \hat{C}_4 and \hat{C}_5 compare the average of \bar{y}_2 , \bar{y}_3 , and \bar{y}_4 against each of the other two sample means, and only \bar{y}_5 is significantly different (less). Finally, \hat{C}_6 compares \bar{y}_5 with the average of the other four class means, and \bar{y}_5 is significantly less. These results are consistent with the results of Duncan's multiple range test in Example 5.7.

Note, however, that if the contrasts had been tested by using the t -test in Equation 5.54, all contrasts except \hat{C}_2 would have been declared significantly larger than zero. This is because $t_{0.95}(14) = 1.76$, whereas the Scheffé criterion is 3.10.

5.2 ONE-WAY RANDOM EFFECTS ANALYSIS OF VARIANCE

In Section 5.1, the effects $\alpha_i = \mu_i - \mu$, for $i = 1, 2, \dots, a$, are called fixed effects because only the a classes present in the experiment are of interest. If the experiment was repeated several times, the same a classes would be included each time. In the fixed effects case, the purpose of the experiment and subsequent data analysis is to estimate and compare the a means μ_i , or the a effects α_i . In some applications, however, the a classes present in an experiment are a random sample from a large population of classes. Thus, if the experiment was repeated several times, a different random sample of a classes would be selected from the population of classes each time. In such cases, the purpose of the experiment, the interpretation of the model used, and the objectives of the analysis are different than for the fixed effects case.

Because the a classes are a random sample from a population of classes, the corresponding true means μ_i are a random sample from a population of means. If it is assumed that the means in this population have a distribution with an overall mean μ , then the effects

$$\alpha_i = \mu_i - \mu, \quad i = 1, 2, \dots, a$$

corresponding to the classes in the experiment, are a random sample from a population of effects that have a distribution with mean zero and variance σ_α^2 . Therefore, the α_i are called **random effects**.

As in the fixed effects case, a random sample of size n_i is taken from the i^{th} class. The measured response for the j^{th} sampled unit in the i^{th} class can be expressed as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (5.63)$$

$$i = 1, 2, \dots, a; j = 1, 2, \dots, n_i$$

where it is assumed that μ is the overall population mean, α_i has a distribution with mean zero and variance σ_α^2 , and ϵ_{ij} has a distribution with mean zero and variance σ^2 . It is also assumed that the α_i are uncorrelated, the ϵ_{ij} are uncorrelated, and every α_i is uncorrelated with every ϵ_{ij} . Under these assumptions, Equation 5.63 is called the **random effects model**.

Although Equations 5.6 and 5.63 have the same form, the interpretation of the α_i and the objectives of the experiment and data analysis differ for the two models. In the fixed effects model, y_{ij} is assumed to have a distribution with mean $\mu_i = \mu + \alpha_i$ and variance σ^2 , and the objectives of the experiment and data analysis are to estimate and compare the μ_i (or α_i), and to estimate σ^2 . Also, any inferences from the fixed effects analysis apply only to the a classes present in the experiment. In the random effects model, however, y_{ij} is assumed to have a distribution with mean μ and variance $\sigma_\alpha^2 + \sigma^2$; that is,

$$E[y_{ij}] = \mu \quad (5.64)$$

and

$$\begin{aligned} \text{Var}[y_{ij}] &= \text{Var}[\alpha_i + \epsilon_{ij}] \\ &= \text{Var}[\alpha_i] + \text{Var}[\epsilon_{ij}] = \sigma_\alpha^2 + \sigma^2 \end{aligned} \quad (5.65)$$

In the random effects case, the objectives of the experiment and data analysis are to estimate μ , σ_α^2 , and σ^2 and to test the hypothesis that $\sigma_\alpha^2 = 0$. Inferences from the random effects analysis are applicable to the entire population of classes, not just the a classes present in the experiment.

5.2.1 Variance Components

For the one-way random effects analysis of variance, the experimental data are tabulated as in Table 5.1. The sums of squares are computed exactly as in the fixed effects case by applying Equations 5.18, 5.19, and 5.20 when the within class sample sizes are equal, or Equations 5.28, 5.30, and 5.32 when the sample sizes are unequal. The degrees of freedom, mean squares, and test statistic F are also computed exactly as in the fixed

effects case. The general form of a one-way random effects analysis of variance table is illustrated by Table 5.3, where the quantity n_0 in the $E[MS]$ column is given by

$$n_0 = \frac{1}{a-1} [N - (\sum_{i=1}^a n_i^2)/N] \quad (5.66)$$

TABLE 5.3
One-Way Random Effects Analysis of Variance Table

Source	df	SS	MS	$E[MS]$	F
Between classes	$a-1$	SS_{Between}	MS_{Between}	$\sigma^2 + n_0\sigma_\alpha^2$	$\frac{MS_{\text{Between}}}{MS_{\text{Within}}}$
Within classes	$N-a$	SS_{Within}	MS_{Within}	σ^2	
Total	$N-1$	SS_{Total}			

When the sample sizes are equal, say $n_i = n$, Equation 5.66 reduces to $n_0 = n$; that is,

$$n_0 = \frac{1}{a-1} [an - an^2/an] = \frac{(a-1)n}{a-1} = n$$

Estimation of the variance components, σ_α^2 and σ^2 , is accomplished by equating the observed and expected mean squares, which results in the equations

$$MS_{\text{Between}} = \hat{\sigma}^2 + n_0 \hat{\sigma}_\alpha^2$$

and

$$MS_{\text{Within}} = \hat{\sigma}^2 \quad (5.67)$$

The estimators of σ^2 and σ_α^2 , obtained by solving these equations, are given by

$$\hat{\sigma}^2 = MS_{\text{Within}} \quad (5.68)$$

and

$$\hat{\sigma}_\alpha^2 = \frac{MS_{\text{Between}} - MS_{\text{Within}}}{n_0} \quad (5.69)$$

These estimators are unbiased when the model, given by Equation 5.63, and the corresponding assumptions are valid. Note that for the purpose of estimation, it is not necessary to assume any particular form of the distribution of y_{ij} except that the mean and variance are given by Equations 5.64 and 5.65.

In practice, it sometimes happens that $MS_{\text{Between}} < MS_{\text{Within}}$, so that Equation 5.69 yields a negative estimate of σ_α^2 . Because a variance is positive by definition, a negative estimate is disturbing. One course of action is to interpret the estimate as evidence that the true value of σ_α^2 is near zero, and that variability due to sampling has caused the negative estimate. Under this philosophy, $\hat{\sigma}_\alpha^2$ is set to zero whenever $MS_{\text{Between}} \leq MS_{\text{Within}}$.

The hypothesis of interest in a one-way random effects analysis is that $\sigma_\alpha^2 = 0$. Because $\sigma_\alpha^2 \geq 0$ by definition, the alternative hypothesis is one-sided and includes values of σ_α^2 larger than 0. That is, we wish to test the hypothesis

$$H_0: \sigma_\alpha^2 = 0$$

against the alternative

$$H_A: \sigma_\alpha^2 > 0$$

(5.70)

at the α level of significance. To do so requires the additional assumptions that

$$\alpha_i \sim N(0, \sigma_\alpha^2) \quad (5.71)$$

$$\epsilon_{ij} \sim N(0, \sigma^2) \quad (5.72)$$

and thus

$$y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2) \quad (5.73)$$

With these additional normality assumptions, $SS_{\text{Within}}/\sigma^2$ has a chi-square distribution with $N - a$ degrees of freedom and $SS_{\text{Between}}/(\sigma_\alpha^2 + n_0\sigma^2)$ has a chi-square distribution with $a - 1$ degrees of freedom. Also, SS_{Between} and SS_{Within} are independent.

From the $E[MS]$ column of Table 5.3, note that if $\sigma_\alpha^2 = 0$, then both MS_{Between} and MS_{Within} are unbiased estimators of σ^2 . Thus, when H_0 is true, the test statistic

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

has an F-distribution with $a - 1$ and $N - a$ degrees of freedom. When $\sigma_\alpha^2 > 0$, however, the numerator in F is an unbiased estimator of $\sigma^2 + n_0\sigma_\alpha^2$, which is larger than σ^2 . That is, when H_A is true, the numerator of F is expected to be larger than the denominator. Thus, H_0 is rejected at the α level of significance if

$$F \geq F_{1-\alpha}(a-1, N-a) \quad (5.74)$$

where $F_{1-\alpha}(a-1, N-a)$ is obtained from Table A6.

Example 5.9 In this example, adopted (and condensed) from Jaech (1973), six $\text{UO}_2\text{-PuO}_2$ pellets are randomly selected from a batch of pellets, and three analyses are performed on each pellet to determine percent plutonium. The resulting data are given in the table below.

Percent Plutonium in $\text{UO}_2\text{-PuO}_2$ Pellets						
	Pellet					
	1	2	3	4	5	6
	2.51	2.46	2.47	2.49	2.51	2.45
	2.53	2.50	2.50	2.53	2.49	2.42
	2.50	2.46	2.47	2.53	2.53	2.50
$y_{i.}$	7.54	7.42	7.44	7.55	7.53	7.37
$y_{..}$						44.85

The population of "classes" includes all pellets in the batch, and the six pellets in the experiment are a random sample from this population. The three analyses on the i^{th} pellet are assumed to be a random sample from

the conceptual population of possible analyses that would be observed if the pellet were analyzed an infinitely large number of times.

The hypothesis that $\sigma_\alpha^2 = 0$ is to be tested at the 0.10 level of significance. It is also of interest to estimate the variance components, σ_α^2 and σ^2 , for pellets and analyses within pellets, respectively, and the variance of y_{ij} .

By applying Equations 5.18, 5.19, and 5.20, the sums of squares are computed as

$$\begin{aligned} SS_{\text{Total}} &= 2.51^2 + 2.53^2 + \dots + 2.50^2 - (44.85)^2 / 18 \\ &= 111.76790 - 111.75125 = 0.01665 \end{aligned}$$

$$\begin{aligned} SS_{\text{Between}} &= (7.54^2 + 7.42^2 + \dots + 7.37^2) / 3 - (44.85)^2 / 18 \\ &= 111.76063 - 111.75125 = 0.00938 \end{aligned}$$

$$SS_{\text{Within}} = 0.01665 - 0.00938 = 0.00727$$

and the analysis of variance table is given below.

Source	df	SS	MS	E[MS]	F
Between pellets	5	0.00938	0.001876	$\sigma^2 + 3\sigma_\alpha^2$	3.10
Within pellets	12	0.00727	0.000606	σ^2	
Total	17	0.01665			

From Table A6, $F_{0.90}(5,12) = 2.39$. Since $3.10 > 2.39$, reject the hypothesis that $\sigma_\alpha^2 = 0$ and conclude that $\sigma_\alpha^2 > 0$.

From Equations 5.68 and 5.69, estimates of σ^2 and σ_α^2 are

$$\hat{\sigma}^2 = 0.000606$$

and

$$\hat{\sigma}_\alpha^2 = \frac{0.001876 - 0.000606}{3} = 0.000423$$

By substituting these estimates in place of σ^2 and σ_α^2 in Equation 5.65, an estimate of the variance of y_{ij} is given by

$$\begin{aligned} \text{V}\hat{\text{a}}\text{r}(y_{ij}) &= \hat{\sigma}_\alpha^2 + \hat{\sigma}^2 \\ &= 0.000423 + 0.000606 = 0.001029 \end{aligned}$$

5.2.2 Power and Required Sample Size for Testing $H_0: \sigma_\alpha^2 = 0$

In some applications, the variance component σ_α^2 is of primary interest, and it is necessary to design a sampling plan that ensures a specified level of power for testing the hypothesis in Equation 5.70. This requires methods for computing the power of the test and for determining the sample size required to ensure a specified level of power.

As a basis for developing the methods presented in this section, note that the ratio

$$\frac{MS_{\text{Between}}/(\sigma^2 + n_0 \sigma_\alpha^2)}{MS_{\text{Within}}/\sigma^2}$$

has an F-distribution with $a - 1$ and $N - a$ degrees of freedom. Thus, the test statistic

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

is distributed as the product

$$U \left(\frac{\sigma^2 + n_0 \sigma_\alpha^2}{\sigma^2} \right)^{-1} = U \left(1 + \frac{n_0 \sigma_\alpha^2}{\sigma^2} \right)^{-1} \quad (5.75)$$

where U has an F-distribution with $a - 1$ and $N - a$ degrees of freedom. To use this result in determining power and required sample sizes for various values of σ_α^2 , a value of σ^2 must be specified. This can be a problem, because σ^2 is usually unknown. A common approach is to work in terms of the quantity

$$\lambda = \sigma_\alpha^2 / \sigma^2 \quad (5.76)$$

For specified values of λ , n_0 , and significance level α (not to be confused with the effects α_i), the power of the test is given by

$$1 - \beta = \Pr[F > (1 + n_0 \lambda)^{-1} F_{1-\alpha}(a-1, N-a)] \quad (5.77)$$

This can be written in the form

$$1 - \beta = \Pr[F > F_{\beta}(a-1, N-a)]$$

where

$$F_{\beta}(a-1, N-a) = (1 + n_0\lambda)^{-1} F_{1-\alpha}(a-1, N-a) \quad (5.78)$$

and where $F_{1-\alpha}(a-1, N-a)$ is obtained from Table A6. Then β (and thus $1-\beta$) is approximated from Table A6 by interpolation. This procedure is illustrated by the following example.

Example 5.10 In Example 5.9, determine the power of the hypothesis test when $\lambda = \sigma_a^2/\sigma^2 = 1$ and when $\lambda = 2$. Recall that $n_0 = n = 3$ and $F_{0.90}(5, 12) = 2.39$.

For the case where $\lambda = 1$, application of Equation 5.78 yields

$$F_{\beta}(5, 12) = (1 + 3)^{-1}(2.39) = 0.598$$

Entering Table A6 with 5 and 12 degrees of freedom, note that all entries in the body of the table are larger than 0.598. Thus, the reciprocal relationship given by Equation 4.54 is used. That is,

$$F_{1-\beta}(12, 5) = 1/F_{\beta}(5, 12)$$

or

$$F_{1-\beta}(12, 5) = 1/0.598 = 1.672$$

Entering Table A6 with 12 and 5 degrees of freedom, the values

$$F_{0.50}(12, 5) = 1.09$$

and

$$F_{0.75}(12, 5) = 1.89$$

are obtained. By linear interpolation, the power of the test when $\lambda = 1$ is approximately

$$1 - \beta \cong 0.75 - \left(\frac{1.89 - 1.672}{1.89 - 1.09} \right) (0.75 - 0.50) = 0.682$$

For the case where $\lambda = 2$, application of Equation 5.78 yields

$$F_{\beta}(5,12) = (1 + 6)^{-1} (2.39) = 0.341$$

and the reciprocal relationship gives

$$F_{1-\beta}(12,5) = 1/0.341 = 2.933$$

Entering Table A6 with 12 and 5 degrees of freedom, the values

$$F_{0.75}(12,5) = 1.89$$

and

$$F_{0.90}(12,5) = 3.27$$

are obtained. By linear interpolation, the power of the test when $\lambda = 2$ is approximately

$$1 - \beta \cong 0.90 - \left(\frac{3.27 - 2.933}{3.27 - 1.89} \right) (0.90 - 0.75) = 0.863$$

Now consider the problem of determining the within class sample size n required to ensure a specified level of power of the hypothesis test when σ_{α}^2 is greater than or equal to a given value. The method presented assumes a fixed number of classes a , and specified values of σ_{α}^2 , σ^2 , significance level α , and power $1 - \beta$. Since σ^2 is usually unknown, it is convenient to work in terms of $\lambda = \sigma_{\alpha}^2/\sigma^2$.

Note that if all $n_i = n$, then $n_0 = n$, and Equation 5.78 can be rearranged and written in the form

$$n = \frac{\left(\frac{F_{1-\alpha}(a-1, N-a)}{F_{\beta}(a-1, N-a)} - 1 \right)}{\lambda} \quad (5.79)$$

Because $N = an$, however, Equation 5.79 involves n nonlinearly and must be solved iteratively for n . The iterations begin by substituting $F_{1-\alpha}(a-1, \infty)$ and $F_{\beta}(a-1, \infty)$, obtained from Table A6, into Equation 5.79 and computing an initial value for n . This first approximation to n (rounded up to the next integer) is used to compute $N = an$, and the quantities $F_{1-\alpha}(a-1, N-a)$ and $F_{\beta}(a-1, N-a)$, obtained from Table A6, are substituted into Equation 5.79. The resulting value of n (rounded up) is then used to compute $N = an$, and the procedure continues until there is no change in the value of n . When the iterations require a value from the F-distribution with degrees of freedom that are not included in Table A6, linear interpolation can be used to compute an approximate F value. Linear interpolation is illustrated in Examples 5.7 and 5.8.

Although the number of classes a is specified in the above procedure, the procedure can be repeated with different values of a . This will determine a number of alternative sampling plans and may be advantageous when the costs of sampling and measurements or analyses are an important consideration.

Example 5.11 The results of Example 5.10 indicate that the sampling plan described in Example 5.9 provides power $1 - \beta = 0.682$ when $\lambda = 1$. Find the required number n of analyses per pellet to ensure that the power is $1 - \beta \geq 0.90$ when $\lambda = 1$, $\alpha = 0.10$, and the number of pellets is $a = 6$.

The iterations begin by obtaining the values of $F_{0.90}(5, \infty) = 1.85$ and $F_{0.10}(5, \infty) = 1/F_{0.90}(\infty, 5) = 1/3.10 = 0.323$ from Table A6. The initial value of n , computed from Equation 5.79, is

$$n = \frac{\left(\frac{1.85}{0.323} - 1 \right)}{1} = 4.73 \cong 5$$

With $N = 6(5) = 30$, the values $F_{0.90}(5, 24) = 2.10$ and $F_{0.10}(5, 24) = 1/F_{0.90}(24, 5) = 1/3.19 = 0.313$ from Table A6 are substituted into Equation 5.79 to compute

$$n = \frac{\left(\frac{2.10}{0.313} - 1 \right)}{1} = 5.71 \cong 6$$

Then with $N = 6(6) = 36$, the values $F_{0.90}(5,30) = 2.05$ and $F_{0.10}(5,30) = 1/F_{0.90}(30,5) = 1/3.17 = 0.315$ from Table A6 are substituted into Equation 5.79, to compute

$$n = \frac{\left(\frac{2.05}{0.315} - 1 \right)}{1} = 5.51 \cong 6$$

Because $n = 6$ is the same value as obtained from the previous iteration, the iterations stop.

Thus, the number of analyses per pellet must be increased from $n = 3$ to $n = 6$ to ensure that the power is at least 0.90 when $\lambda = 1$ (i.e., when $\sigma_{\alpha}^2 = \sigma^2$).

5.2.3 Estimation of the Overall Mean μ

An important objective of a one-way random effects analysis of variance is often to estimate the overall mean μ . When the n_i are equal, say $n_i = n$, the minimum variance unbiased estimator of μ is given by

$$\bar{y}_{..} = y_{..}/N \quad (5.80)$$

When the n_i are unequal, however, Equation 5.80 assigns more weight to classes with larger n_i values and less weight to classes with smaller n_i values. To see this, note that

$$\bar{y}_{..} = \frac{1}{N} y_{..} = \frac{1}{N} \sum_{i=1}^a y_i = \frac{1}{N} \sum_{i=1}^a n_i \bar{y}_i$$

where $\bar{y}_i = y_i/n_i$ is the sample mean of the i^{th} class. If the unequal n_i were planned in advance for the specific purpose of weighting the class sample means in estimating μ , then Equation 5.80 may be an appropriate estimator. With unequal n_i , however, although $\bar{y}_{..}$ is an unbiased estimator for μ , it is not a minimum variance estimator unless $\sigma_{\alpha}^2 = 0$. Thus, for

applications where $\sigma_\alpha^2 > 0$, the investigator should carefully consider the trade-off between achieving a particular weighting of the observations in estimating μ and obtaining a more precise estimate (one with a smaller variance).

In practice, the unequal n_i are usually not planned, but happen due to unavoidable occurrences or circumstances during experimentation or measurement. In such cases, it is desirable to find an unbiased estimator for μ that has the smallest possible variance. Searle and Pukelsheim (1986) show that the minimum variance unbiased estimator for μ is the weighted average of the class sample means, given by

$$\hat{\mu} = \frac{\sum_{i=1}^a \frac{n_i \bar{y}_i}{(n_i \sigma_\alpha^2 + \sigma^2)}}{\sum_{i=1}^a \frac{n_i}{(n_i \sigma_\alpha^2 + \sigma^2)}}$$

where the weight assigned to each \bar{y}_i is the reciprocal of its variance.

Unfortunately, σ_α^2 and σ^2 are usually unknown, and $\hat{\mu}$ cannot be applied directly. One option is to substitute the estimates $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}^2$. When this is done, however, the variance of $\hat{\mu}$ involves the variances and covariance of $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}^2$, which results in an expression that is cumbersome and difficult to work with. To gain an appreciation for the potential complexity of the variance of $\hat{\mu}$, the interested reader is referred to the formulas for the variances and covariance of $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}^2$ in Searle (1971). Also, when estimates are substituted for σ_α^2 and σ^2 in $\hat{\mu}$, the variance of $\hat{\mu}$ is larger (in many cases considerably larger) than when σ_α^2 and σ^2 are known, and $\hat{\mu}$ is no longer the minimum variance estimator.

In most nuclear material accounting applications, the variance components are unknown, and one objective of the experiment and analysis of variance is to estimate them. Thus, $\hat{\mu}$ is not recommended for general use. Instead, two other estimators are discussed and recommendations are given to help in deciding when each is most appropriate.

The discussion given by Searle and Pukelsheim (1986) makes use of the ratio

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$$

which is referred to in the statistics literature as the intraclass correlation coefficient; e.g., see Snedecor and Cochran (1980). As ρ approaches zero, $\hat{\mu}$ approaches the estimator $\bar{y}_.$, given by Equation 5.80. A value of ρ near

zero indicates that σ_α^2 is very small relative to σ^2 . As ρ approaches one, $\hat{\mu}$ approaches the unweighted average of the class sample means, given by

$$\bar{y}_u = \sum_{i=1}^a \bar{y}_i / a \quad (5.81)$$

A value of ρ near one indicates that σ^2 is very small relative to σ_α^2 .

By letting ρ vary between zero and one, Searle and Pukelsheim (1986) demonstrate that as ρ increases from zero, the weights assigned to the class sample means in $\hat{\mu}$ rapidly equalize (i.e., become nearly equal). The rate of equalization depends upon the values of the n_i . In the most extreme cases considered, however, the weights were nearly equal when $\rho \geq 0.7$, which corresponds to values of $\sigma^2 \leq 0.4\sigma_\alpha^2$. In such cases, $\hat{\mu}$ is practically the same as the unweighted estimator \bar{y}_u , given by Equation 5.81.

For many applications, neither \bar{y}_u nor $\bar{y}_..$ will be a minimum variance estimator. For specified values of the variance components and the n_i , however, one of the estimators will have a smaller variance than the other and would be a satisfactory choice.

It can be shown that when $\sigma_\alpha^2 \geq \sigma^2$ (which is often the case), \bar{y}_u has a smaller variance than $\bar{y}_..$ for any set of unequal n_i . Thus \bar{y}_u is recommended whenever $\hat{\sigma}_\alpha^2 \geq \hat{\sigma}^2$. If $\hat{\sigma}_\alpha^2$ is quite small relative to $\hat{\sigma}^2$, then $\bar{y}_..$ is recommended. For cases that do not satisfy one of the above criteria, it is recommended that the estimated variances of \bar{y}_u and $\bar{y}_..$ be computed (formulas are given below); then the estimator with the smallest estimated variance should be used.

Note that when the n_i are equal, $\bar{y}_.. = \bar{y}_u = \hat{\mu}$. This is one advantage of equal sample sizes.

The variance of $\bar{y}_..$ is given by

$$\text{Var}(\bar{y}_..) = \frac{r_1 \sigma_\alpha^2 + \sigma^2}{N} \quad (5.82)$$

where

$$r_1 = \sum_{i=1}^a n_i^2 / N \quad (5.83)$$

$\text{Var}(\bar{y}_..)$ is estimated by substituting $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}^2$, computed from Equations 5.68 and 5.69, into Equation 5.82, which gives

$$\text{Vâr}(\bar{y}_{..}) = \frac{r_1 \hat{\sigma}_\alpha^2 + \hat{\sigma}^2}{N} \quad (5.84)$$

The degrees of freedom for $\text{Vâr}(\bar{y}_{..})$ are approximated by applying Satterthwaite's (1946) formula for linear combinations of independent mean squares. This requires that Equation 5.84 be expressed in terms of MS_{Between} and MS_{Within} . From Equations 5.68 and 5.69,

$$\begin{aligned} \text{Vâr}(\bar{y}_{..}) &= \frac{\frac{r_1 \left(MS_{\text{Between}} - MS_{\text{Within}} \right)}{n_0} + MS_{\text{Within}}}{N} \\ &= \frac{r_1}{N n_0} MS_{\text{Between}} + \left(\frac{1}{N} - \frac{r_1}{N n_0} \right) MS_{\text{Within}} \\ &= m_1 MS_{\text{Between}} + m_2 MS_{\text{Within}} \end{aligned} \quad (5.85)$$

The approximate degrees of freedom for $\text{Vâr}(\bar{y}_{..})$ are then

$$\hat{d}f = \frac{(m_1 MS_{\text{Between}} + m_2 MS_{\text{Within}})^2}{\frac{(m_1 MS_{\text{Between}})^2}{a-1} + \frac{(m_2 MS_{\text{Within}})^2}{N-a}} \quad (5.86)$$

and $\hat{d}f$ is rounded to the nearest integer. When it is desirable to be conservative, $\hat{d}f$ is rounded down.

The variance of the unweighted estimator \bar{y}_u is given by

$$\text{Var}(\bar{y}_u) = \frac{\sigma_\alpha^2 + r_2 \sigma^2}{a} \quad (5.87)$$

where

$$r_2 = \frac{1}{a} \sum_{i=1}^a \left(\frac{1}{n_i} \right) \quad (5.88)$$

An estimate of $\text{Var}(\bar{y}_u)$ is given by

$$\hat{\text{Vâr}}(\bar{y}_u) = \frac{\hat{\sigma}_\alpha^2 + r_2 \hat{\sigma}^2}{a} \quad (5.89)$$

In terms of the mean squares, $\hat{\text{Vâr}}(\bar{y}_u)$ can be expressed as

$$\begin{aligned} \hat{\text{Vâr}}(\bar{y}_u) &= \frac{\left(\frac{\text{MS}_{\text{Between}} - \text{MS}_{\text{Within}}}{n_0} \right) + r_2 \text{MS}_{\text{Within}}}{a} \\ &= \frac{\text{MS}_{\text{Between}}}{an_0} + \left(\frac{r_2}{a} - \frac{1}{an_0} \right) \text{MS}_{\text{Within}} \\ &= m_3 \text{MS}_{\text{Between}} - m_4 \text{MS}_{\text{Within}} \end{aligned} \quad (5.90)$$

The approximate degrees of freedom for $\hat{\text{Vâr}}(\bar{y}_u)$ are then

$$\hat{d}f_u = \frac{(m_3 \text{MS}_{\text{Between}} + m_4 \text{MS}_{\text{Within}})^2}{\frac{(m_3 \text{MS}_{\text{Between}})^2}{a-1} + \frac{(m_4 \text{MS}_{\text{Within}})^2}{N-a}} \quad (5.91)$$

and $\hat{d}f_u$ is rounded to the nearest integer. When it is desirable to be conservative, $\hat{d}f_u$ is rounded down.

At this point, it should be noted that when the n_i are equal, say $n_i = n$, the estimators $\bar{y}_..$ and \bar{y}_u are identical, and Equations 5.82 and 5.87 reduce to the form

$$\begin{aligned} \text{Var}(\bar{y}_..) = \text{Var}(\bar{y}_u) &= \frac{n\sigma_\alpha^2 + \sigma^2}{an} \\ &= \frac{E[\text{MS}_{\text{Between}}]}{an} \end{aligned} \quad (5.92)$$

Then, Equations 5.84 and 5.89 reduce to

$$\hat{\text{Vâr}}(\bar{y}_..) = \hat{\text{Vâr}}(\bar{y}_u) = \frac{\text{MS}_{\text{Between}}}{an} \quad (5.93)$$

and Equations 5.86 and 5.91 reduce to

$$\hat{df} = \hat{df}_u = a - 1 \quad (5.94)$$

Assuming that $y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2)$, an approximate $100(1 - \alpha)\%$ confidence interval for μ is given by either

$$\bar{y}_{..} \pm \sqrt{\text{Vâr}(\bar{y}_{..})} t_{1-\alpha/2}(\hat{df}) \quad (5.95)$$

or

$$\bar{y}_u \pm \sqrt{\text{Vâr}(\bar{y}_u)} t_{1-\alpha/2}(\hat{df}_u) \quad (5.96)$$

depending upon whether $\bar{y}_{..}$ or \bar{y}_u is chosen to estimate μ . In Equations 5.95 and 5.96, the value of $t_{1-\alpha/2}(\hat{df})$ or $t_{1-\alpha/2}(\hat{df}_u)$ is obtained from Table A5.

Hypotheses about μ are tested by using methods analogous to those in Section 4.2.2. The test statistic is given by either

$$t = \frac{\bar{y}_{..} - \mu_0}{\sqrt{\text{Vâr}(\bar{y}_{..})}} \quad (5.97)$$

or

$$t = \frac{\bar{y}_u - \mu_0}{\sqrt{\text{Vâr}(\bar{y}_u)}} \quad (5.98)$$

depending upon whether $\bar{y}_{..}$ or \bar{y}_u is chosen to estimate μ . In the notation of Section 4.2.2, μ_0 is a hypothesized value of μ , and the test statistic t is compared with a value obtained from Table A5 with either \hat{df} or \hat{df}_u degrees of freedom.

5.2.4 Sampling to Estimate μ

Suppose that an experiment is to be designed by selecting n and a so that $\text{Var}(\bar{y}_{..})$ in Equation 5.92 is minimized, subject to limits on the cost of experimentation. In many applications, the cost C of randomly selecting a units from a population and making n independent measurements or analyses on each unit is approximately of the form

$$C = c_1a + c_2an \quad (5.99)$$

In this cost function, c_1 is the cost of an individual unit and its random selection from the population, and c_2 is the cost of a single measurement or analysis on a unit or on a subsample from the unit. It can be shown that $\text{Var}(\bar{y}_{..})$ is minimized with respect to a and n , subject to a fixed total cost C , when

$$n = \sqrt{c_1 \sigma^2 / c_2 \sigma_\alpha^2} \quad (5.100)$$

Because n must be a positive integer, this value is rounded to the nearest positive integer, and a is found from Equation 5.99.

Use of Equation 5.100 requires values for σ_α^2 and σ^2 , which are usually unknown. Thus, if an extensive experiment is being planned to estimate the population mean μ , a small pilot study may be needed to obtain preliminary estimates of σ_α^2 and σ^2 .

Example 5.12 In Example 5.9, six $\text{UO}_2\text{-PuO}_2$ pellets are randomly selected from a batch of pellets, and three analyses are performed on each pellet to determine percent plutonium. The following estimates are obtained from the sample data in Example 5.9:

$$\bar{y}_{..} = y_{..}/N = 44.85/18 = 2.4917$$

$$\hat{\sigma}_\alpha^2 = 0.000423$$

$$\hat{\sigma}^2 = 0.000606$$

$$MS_{\text{Between}} = 0.001876$$

Suppose that performing each analysis costs five times as much as the combined cost of randomly selecting a pellet and the loss of some material due to analyzing the pellet (most of the pellet would reenter the manufacturing process as usable scrap). Can the experiment be redesigned to reduce $\text{Var}(\bar{y}_{..})$ while holding the cost of the experiment fixed?

First, let c_2 be the cost of each analysis on a pellet. Thus, $c_1 = c_2/5$ is the combined cost of randomly selecting a pellet and the loss of some material. From Equation 5.99, the total cost of the experiment described in Example 5.9 is

$$C = 6c_1 + 18c_2 = (6c_2/5) + 18c_2 = 19.2c_2$$

An estimate of $\text{Var}(\bar{y}_{..})$, computed from Equation 5.93, is

$$\hat{\text{Var}}(\bar{y}_{..}) = \frac{0.001876}{18} = 0.0001042$$

If the estimates of σ_α^2 and σ^2 are substituted into Equation 5.100 with $c_1 = c_2/5$, the resulting value of n is

$$\begin{aligned} n &= \sqrt{(c_2/5)(0.000606)/c_2(0.000423)} \\ &= \sqrt{0.000606/5(0.000423)} = 0.535 \end{aligned}$$

Because n must be a positive integer, $n = 1$ is chosen. Thus, $\text{Var}(\bar{y}_{..})$ is minimized, subject to a fixed total cost, by performing only one analysis on each pellet. To determine a , the number of pellets to include in the experiment, use the cost function in Equation 5.99 with $n = 1$. This gives

$$19.2c_2 = \frac{c_2(a)}{5} + c_2(a)(1) = 1.2(c_2)(a)$$

or

$$a = \frac{19.2}{1.2} = 16 \text{ pellets}$$

By substituting the sample estimates of σ_α^2 and σ^2 with $a = 16$ and $n = 1$ into Equation 5.92, an estimate of the resulting variance of $\bar{y}_{..}$ is

$$\frac{1(0.000423) + 0.000606}{16(1)} = 0.0000643$$

For comparison of this value with the estimate $\hat{\text{Var}}(\bar{y}_{..}) = 0.0001042$ from the experiment described in Example 5.9 the ratio is

$$\frac{0.0000643}{0.0001042} = 0.62$$

That is, $\text{Var}(\bar{y}_{..})$ for the revised experiment is expected to be about 62% of that for the previous experiment. These results indicate that for the same cost as the experiment described in Example 5.9, a more precise estimate

of μ (one with a smaller variance) could be obtained by performing only one analysis on each of 16 randomly selected pellets.

Note, however, that with $n = 1$ analysis per pellet, it will not be possible to compute MS_{within} , to separately estimate σ_α^2 and σ^2 , or to test the hypothesis that $\sigma_\alpha^2 = 0$. This may not be a problem if the primary objective is to estimate μ . But if it is necessary to separately estimate σ_α^2 and σ^2 (for example, if σ_α^2 is being monitored), then it is necessary to have some $n_i > 1$. The values of the n_i must be chosen to meet the objectives of the experiment to the extent possible, subject to the cost limitations. Thus, there could be trade-offs between, for example, minimizing $\text{Var}(\bar{y}_{..})$, attaining a specified level of power in testing σ_α^2 , and meeting the cost constraints. In fact, it might be necessary to evaluate a number of alternative sampling plans and then select the one that is the best compromise.

5.3 ANALYSIS OF VARIANCE WITH NESTED (HIERARCHICAL) MODELS

The concepts presented in Section 5.1 and 5.2 can be extended to three or more stages of sampling. To illustrate this, suppose that the sampling plan described in Example 5.9 is carried out on each of four batches of pellets from a process. That is, six pellets are randomly selected from each of four batches, and three analyses are performed on each pellet to determine percent plutonium. If y_{ijk} denotes the k^{th} analysis on the j^{th} pellet from the i^{th} batch, then a linear model that characterizes y_{ijk} is given by

$$y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + \epsilon_{(ij)k} \quad (5.101)$$

$$i = 1, \dots, a; j = 1, \dots, b_i; k = 1, \dots, n_{ij}$$

The terms in this model are defined as follows:

- μ = the overall population mean percent plutonium
- α_i = the deviation from μ due to the i^{th} batch. In context of Sections 5.1 and 5.2, α_i is the effect of the i^{th} batch. There are $a = 4$ batches in this example.
- $\beta_{(i)j}$ = the deviation from $\mu + \alpha_i$ due to the j^{th} pellet in the i^{th} batch. There are $b_i = 6$ pellets from batch i in this example.
- $\epsilon_{(ij)k}$ = the deviation from $\mu + \alpha_i + \beta_{(i)j}$ due to the k^{th} analysis of the ij^{th} pellet. There are $n_{ij} = 3$ analyses on the ij^{th} pellet in this example.

In this section, analysis of variance procedures arising from two different

assumptions about α_i are presented. In the first case, the α_i , $\beta_{(i)j}$, and $\epsilon_{(ij)k}$ are all assumed to be random effects. The resulting model is called the nested (or hierarchical) random effects model. In the second case, the α_i are assumed to be fixed effects, with $\beta_{(i)j}$ and $\epsilon_{(ij)k}$ assumed to be random effects. The resulting model is called the nested (or hierarchical) mixed effects model. The term "mixed effects model," or simply "mixed model," is used in the analysis of variance literature to indicate that a model includes both fixed and random effects.

5.3.1 The Nested Random Effects Model

Suppose that in the example described in the previous paragraph, the four batches of pellets are assumed to be a random sample from a large number of batches from the process over a given time period. Assume also that a random sample of six pellets is taken from each batch and that the three analyses performed on a given pellet are a random sample from the population of all possible analyses of the pellet. The appropriate model is given by Equation 5.101, where the terms in the model are defined as follows:

- μ = mean percent plutonium over all pellets in all batches
- α_i = the deviation from μ due to the i^{th} batch; α_i is a random observation from a population with zero mean and with variance σ_α^2
- $\beta_{(i)j}$ = the deviation from $\mu + \alpha_i$ due to the ij^{th} pellet; $\beta_{(i)j}$ is a random observation from a population with zero mean and with variance σ_β^2
- $\epsilon_{(ij)k}$ = the deviation from $\mu + \alpha_i + \beta_{(i)j}$ due to the k^{th} analysis of the ij^{th} pellet; $\epsilon_{(ij)k}$ is a random observation from a population with zero mean and with variance σ^2 .

It is also assumed that the α_i are uncorrelated, the $\beta_{(i)j}$ are uncorrelated, the $\epsilon_{(ij)k}$ are uncorrelated, and there are no correlations among the α_i , $\beta_{(i)j}$, and $\epsilon_{(ij)k}$.

Thus, y_{ijk} is an observation from a population with mean μ and with variance $\sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$. In general, it is of interest to

1. Estimate σ_α^2 , σ_β^2 , and σ^2 .
2. Test the hypotheses $H_0: \sigma_\alpha^2 = 0$ and $H_0: \sigma_\beta^2 = 0$.
3. Estimate μ .

5.3.1.1 Variance Components: Estimation and Hypothesis Tests

The two sets of hypotheses that are of interest for the nested random effects model are

$$\begin{aligned} 1. \quad H_0: \sigma_\alpha^2 &= 0 \\ H_A: \sigma_\alpha^2 &> 0 \end{aligned} \tag{5.102}$$

and

$$\begin{aligned} 2. \quad H_0: \sigma_\beta^2 &= 0 \\ H_A: \sigma_\beta^2 &> 0 \end{aligned} \tag{5.103}$$

The quantities that must be computed to test these hypotheses and to obtain estimates of the variance components σ_α^2 , σ_β^2 , and σ^2 are summarized as follows:

$$y_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk} \quad y_{i.} = \sum_{j=1}^{b_i} y_{ij} \quad y_{...} = \sum_{i=1}^a y_{i.} \tag{5.104}$$

$$n_{i.} = \sum_{j=1}^{b_i} n_{ij} \quad n_{..} = \sum_{i=1}^a n_{i.} \quad b_{..} = \sum_{i=1}^a b_i \tag{5.105}$$

$$SS_\alpha = \sum_{i=1}^a (y_{i.}^2/n_{i.}) - y_{...}^2/n_{..} \tag{5.106}$$

$$SS_\beta = \sum_{i=1}^a \sum_{j=1}^{b_i} (y_{ij}^2/n_{ij}) - \sum_{i=1}^a (y_{i.}^2/n_{i.}) \tag{5.107}$$

$$SS_\epsilon = \sum_{i=1}^a \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^a \sum_{j=1}^{b_i} (y_{ij}^2/n_{ij}) \tag{5.108}$$

$$df_\alpha = a - 1 \tag{5.109}$$

$$df_\beta = b_{..} - a \tag{5.110}$$

$$df_\epsilon = n_{..} - b_{..} \tag{5.111}$$

$$MS_\alpha = SS_\alpha/df_\alpha \tag{5.112}$$

$$MS_{\beta} = SS_{\beta}/df_{\beta} \quad (5.113)$$

$$MS_{\epsilon} = SS_{\epsilon}/df_{\epsilon} \quad (5.114)$$

$$k_1 = \frac{\left(\sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}^2/n_i \right) - \left(\sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}^2/n_{..} \right)}{a - 1} \quad (5.115)$$

$$k_2 = \frac{n_{..} - \sum_{i=1}^a n_i^2/n_{..}}{a - 1} \quad (5.116)$$

$$k_3 = \frac{n_{..} - \sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}^2/n_i}{b_{.} - a} \quad (5.117)$$

The values computed from Equations 5.106 through 5.117 are then displayed in an analysis of variance table as illustrated by Table 5.4.

TABLE 5.4
Nested Random Effects Analysis of Variance Table

Source	df	SS	MS	E[MS]
α_i	$a - 1$	SS_{α}	MS_{α}	$\sigma^2 + k_1\sigma_{\beta}^2 + k_2\sigma_{\alpha}^2$
$\beta_{(i)j}$	$b_{.} - a$	SS_{β}	MS_{β}	$\sigma^2 + k_3\sigma_{\beta}^2$
$\epsilon_{(ij)k}$	$n_{..} - b_{.}$	SS_{ϵ}	MS_{ϵ}	σ^2

Estimates of the variance components σ_{α}^2 , σ_{β}^2 , and σ^2 are obtained by equating the computed mean squares to their expected values and then solving the resulting equations for the unknown variance components. This leads to the unbiased estimators

$$\hat{\sigma}^2 = MS_{\epsilon} \quad (5.118)$$

$$\hat{\sigma}_{\beta}^2 = \frac{MS_{\beta} - MS_{\epsilon}}{k_3} \quad (5.119)$$

and

$$\begin{aligned}\hat{\sigma}_\alpha^2 &= \frac{MS_\alpha - k_1 \hat{\sigma}_\beta^2 - MS_\epsilon}{k_2} \\ &= \frac{MS_\alpha - \frac{k_1}{k_3} (MS_\beta - MS_\epsilon) - MS_\epsilon}{k_2}\end{aligned}\quad (5.120)$$

If $\hat{\sigma}_\beta^2$ yields a negative estimate, then the estimate is set equal to zero, and zero is substituted for $\hat{\sigma}_\beta^2$ in Equation 5.120. When this is done, however, the resulting estimates of $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_\alpha^2$ are not unbiased. Similarly, if $\hat{\sigma}_\alpha^2$ yields a negative estimate, the estimate is set equal to zero and is no longer an unbiased estimate of σ_α^2 .

The test of the hypothesis in Equation 5.103 is carried out by computing the test statistic

$$F = MS_\beta / MS_\epsilon \quad (5.121)$$

Under the null hypothesis, $H_0: \sigma_\beta^2 = 0$, and the assumption that y_{ijk} is from a normal distribution, the test statistic F has an F -distribution with degrees of freedom df_β and df_ϵ . Thus, H_0 is rejected at the α level of significance if

$$F \geq F_{1-\alpha}(b, -a, n_{..} - b) \quad (5.122)$$

where $F_{1-\alpha}(b, -a, n_{..} - b)$ is obtained from Table A6.

In general, unless $k_1 = k_3$, it is necessary to perform an approximate test of the hypothesis given by Equation 5.102. The approximate test presented here is from Satterthwaite (1946) and requires computing the test statistic

$$F^* = MS_\alpha / MS_\beta^* \quad (5.123)$$

where

$$MS_\beta^* = \frac{k_1}{k_3} MS_\beta + \left(1 - \frac{k_1}{k_3}\right) MS_\epsilon = m_1 MS_\beta + m_2 MS_\epsilon \quad (5.124)$$

The null hypothesis, $H_0: \sigma_\alpha^2 = 0$, is rejected at approximately the α level of significance (not to be confused with the effects α_i) if

$$F^* \geq F_{1-\alpha}(a-1, df_\beta^*) \quad (5.125)$$

where $F_{1-\alpha}(a-1, df_\beta^*)$ is obtained from Table A6, and where

$$df_\beta^* = \frac{(m_1 MS_\beta + m_2 MS_\epsilon)^2}{\frac{(m_1 MS_\beta)^2}{(b-a)} + \frac{(m_2 MS_\epsilon)^2}{(n_{..} - b)}} \quad (5.126)$$

This quantity is rounded to the nearest integer. When it is desirable to be conservative, df_β^* is rounded down.

In the approximate test, MS_β^* has expected value

$$E[MS_\beta^*] = \sigma^2 + k_1 \sigma_\beta^2 \quad (5.127)$$

Thus, if y_{ijk} is from a normal distribution, the test statistic F^* , given by Equation 5.123, is approximately distributed as the product

$$U \left(\frac{\sigma^2 + k_1 \sigma_\beta^2 + k_2 \sigma_\alpha^2}{\sigma^2 + k_1 \sigma_\beta^2} \right) \quad (5.128)$$

where U has an F-distribution with $a - 1$ and df_β^* degrees of freedom. When $\sigma_\alpha^2 = 0$, Equation 5.128 reduces to U , and the test statistic F^* has approximately an F-distribution with $a - 1$ and df_β^* degrees of freedom. However, when $\sigma_\alpha^2 > 0$, Equation 5.128 yields a value larger than U . Thus, a value of F^* which is greater than or equal to $F_{1-\alpha}(a-1, df_\beta^*)$ indicates that σ_α^2 is probably larger than zero.

Example 5.13 Random samples from a single batch of Pu oxide are distributed to each of three laboratories for analysis of percent plutonium by a given analytic method. These three laboratories are assumed to be a random sample from the population of possible laboratories that could have been asked to participate in this study. Within each laboratory, a number of aliquots are prepared in such a way that, for a given laboratory, they are assumed to be a random sample from all possible aliquots that could have been prepared by that laboratory. Then replicate analyses are performed on each aliquot within each laboratory, and it is assumed that the analyses on a given aliquot are a random sample from the population of all possible analyses of the aliquot. The resulting data, in percent recov-

ery minus 86, are tabulated below with quantities computed from Equations 5.104 and 5.105.

Lab Aliquot	1			2			3	
	1	2	3	1	2	3	1	2
	0.188	0.348	0.388	0.847	1.007	0.837	0.651	0.581
	0.178	0.298	0.338		0.837	0.657	0.611	0.641
	0.158	0.368	0.398				0.591	
	0.198							
y_{ij}	0.722	1.014	1.124	0.847	1.844	1.494	1.853	1.222
$y_{i..}$			2.860			4.185		3.075
$y_{...}$								10.120
n_{ij}	4	3	3	1	2	2	3	2
$n_{i.}$			10			5		5
$n_{..}$								20
b_i			3			3		2
b								8

The objectives of this experiment are

1. To estimate the variance components for laboratories, aliquots within laboratories, and replicate analyses on aliquots within laboratories.
2. To test the hypotheses that the variance components for laboratories and aliquots within laboratories are zero.

The nested random effects analysis of variance with three stages of sampling is the appropriate methodology for this application. Considering the linear model given by Equation 5.101, the terms in the model are defined for this application as follows:

- μ = true mean percent uranium for the population from which samples were randomly selected to send to the three laboratories
- α_i = deviation from μ due to the i^{th} laboratory; α_i is assumed to be a random observation from a population with zero mean and with variance σ_{Labs}^2
- $\beta_{(i)j}$ = deviation from $\mu + \alpha_i$ due to the j^{th} aliquot within the i^{th} laboratory; $\beta_{(i)j}$ is assumed to be a random observation from a population with zero mean and with variance σ_{Aliq}^2
- $\epsilon_{(ij)k}$ = deviation from $\mu + \alpha_i + \beta_{(i)j}$ due to the k^{th} replicate analysis of the j^{th} aliquot within the i^{th} laboratory; $\epsilon_{(ij)k}$ is assumed to be a random observation from a population with zero mean and with variance σ_{Anal}^2

Equations 5.106 through 5.117 are applied to compute the following quantities used to construct the analysis of variance table displayed below.

$$\begin{aligned} SS_{\text{Labs}} &= \left(\frac{2.860^2}{10} + \frac{4.185^2}{5} + \frac{3.075^2}{5} \right) - \frac{10.120^2}{20} \\ &= 6.211930 - 5.120720 = 1.091210 \end{aligned}$$

$$\begin{aligned} SS_{\text{Aliquots}} &= \left(\frac{0.722^2}{4} + \frac{1.014^2}{3} + \frac{1.124^2}{3} + \dots + \frac{1.222^2}{2} \right) \\ &\quad - 6.211930 = 6.318952 - 6.211930 = 0.107022 \end{aligned}$$

$$\begin{aligned} SS_{\text{Analyses}} &= (0.188^2 + 0.178^2 + \dots + 0.641^2) - 6.318952 \\ &= 6.358810 - 6.318952 = 0.039858 \end{aligned}$$

$$df_{\text{Labs}} = 3 - 1 = 2$$

$$df_{\text{Aliq}} = 8 - 3 = 5$$

$$df_{\text{Anal}} = 20 - 8 = 12$$

$$\begin{aligned} k_1 &= \frac{1}{(3-1)} \left(\frac{4^2 + 3^2 + 3^2}{0} + \frac{1^2 + 2^2 + 2^2}{5} \right. \\ &\quad \left. + \frac{3^2 + 2^2}{5} - \frac{4^2 + 3^2 + \dots + 2^2}{20} \right) \\ &= (7.8 - 2.8)/2 = 2.50 \end{aligned}$$

$$k_2 = \frac{20 - (10^2 + 5^2 + 5^2)/20}{3-1} = \frac{20 - (150/20)}{2} = 6.25$$

$$k_3 = \frac{20 - 7.8}{8-3} = 2.44$$

Source	df	SS	MS	E[MS]
Labs	2	1.091210	0.545605	$\sigma_{\text{Anal}}^2 + 2.50 \sigma_{\text{Aliq}}^2 + 6.25 \sigma_{\text{Labs}}^2$
Aliquots	5	0.107022	0.021404	$\sigma_{\text{Anal}}^2 + 2.44 \sigma_{\text{Aliq}}^2$
Analyses	12	0.039858	0.003322	σ_{Anal}^2

The variance component estimates, computed from Equations 5.118, 5.119, and 5.120, are

$$\hat{\sigma}_{\text{Anal}}^2 = 0.003322$$

$$\hat{\sigma}_{\text{Aliq}}^2 = \frac{0.021404 - 0.003322}{2.44} = 0.007411$$

$$\hat{\sigma}_{\text{Labs}}^2 = \frac{0.545605 - 2.50(0.007411) - 0.003322}{6.25} = 0.083801$$

To test the hypothesis

$$H_0: \sigma_{\text{Aliq}}^2 = 0$$

against the alternative

$$H_A: \sigma_{\text{Aliq}}^2 > 0$$

the test statistic F , computed from Equation 5.121, is

$$F = \frac{MS_{\text{Aliq}}}{MS_{\text{Anal}}} = \frac{0.021404}{0.003322} = 6.44$$

From Table A6, the value $F_{0.95}(5,12) = 3.11$ is obtained. Because $6.44 > 3.11$, H_0 is rejected, and it is concluded that σ_{Aliq}^2 is probably larger than zero.

To perform the approximate test of the hypothesis

$$H_0: \sigma_{\text{Labs}}^2 = 0$$

against the alternative

$$H_A: \sigma_{\text{Labs}}^2 > 0$$

Equation 5.124 is used to compute the quantity

$$\begin{aligned} MS_{\text{Aliq}}^* &= \frac{2.50}{2.44} MS_{\text{Aliq}} + \left(1 - \frac{2.50}{2.44}\right) MS_{\text{Anal}} \\ &= 0.021931 - 0.000082 = 0.021849 \end{aligned}$$

which has approximate degrees of freedom computed from Equation 5.126 to be

$$df_{\text{Aliq}}^* = \frac{(0.021849)^2}{\frac{(0.021931)^2}{5} + \frac{(-0.000082)^2}{12}} = 4.96 \cong 5$$

The test statistic F^* , computed from Equation 5.123, is

$$F^* = \frac{MS_{\text{Labs}}}{MS_{\text{Aliq}}^*} = \frac{0.545605}{0.021849} = 24.97$$

From Table A6, the value $F_{0.95}(2,5) = 5.79$ is obtained. Because $24.97 > 5.79$, H_0 is rejected, and it is concluded that the σ_{Labs}^2 is probably larger than zero.

5.3.1.2 Estimation of the Overall Mean μ

When the b_i and n_{ij} are not constant over the experiment, there are a number of important considerations in choosing an appropriate estimator for μ from the nested random effects model. One consideration is whether or not the unequal sample sizes were planned in advance for the specific purpose of assigning more (or less) weight to certain sampling units in estimating μ . If so, the intended weighting scheme will dictate the form of the estimator. Usually, however, the unequal sample sizes were not planned in advance, in which case the objective is to determine an unbiased estimator for μ that has the smallest possible variance.

It is possible to construct a minimum variance unbiased estimator for μ . In Example 5.13, the estimator would be a weighted average of the lab averages, and the lab averages would be weighted averages of the aliquot averages. The weights, however, would be functions of the unknown variance components. This raises the same issues discussed in Section 5.2.3 in connection with the estimator $\hat{\mu}$. Specifically, when estimates are substituted for the unknown variance components, the variance of the estimator for μ is larger (in some cases considerably larger) than when the variance components are known, and the estimator is no longer a minimum variance estimator. Also, when the variance components are unknown, the variance of the estimator for μ involves the variances and covariances of the variance component estimators, which results in an expression that is cumbersome and difficult to use.

Thus, rather than attempting to provide a single estimator for all cases, four estimators for μ are presented in this section with guidelines for deciding which is most appropriate for a particular application. It should be noted that for most applications, none of the four estimators will be a minimum variance estimator. However, for specified values of the variance components, the b_i and the n_{ij} , one of the estimators will have a smaller variance than the others and would be a satisfactory choice.

Table 5.5 lists four possible schemes for using the unequal b_i and n_{ij} as weights in estimating μ from a nested random effects model. The corresponding estimator for each scheme is indicated. It should be noted that when the b_i and n_{ij} are constant over the experiment, all four estimators of μ are equivalent, and the estimation techniques are considerably simplified. See Section 5.3.1.3 for details.

TABLE 5.5
Schemes for Using Unequal b_i and n_{ij} as
Weights in Estimating μ in Example 5.13

Weighting scheme	Weight lab averages equally?	Weight aliquot averages equally?	Estimator
1	Yes	Yes	\bar{y}_{uu}
2	Yes	No	\bar{y}_{uw}
3	No	Yes	\bar{y}_{wu}
4	No	No	\bar{y}_{ww}

Weighting scheme 1 in Table 5.5 leads to the completely unweighted estimator given by

$$\bar{y}_{uu} = \frac{1}{a} \sum_{i=1}^a \frac{1}{b_i} \sum_{j=1}^{b_i} (y_{ij}/n_{ij}) \quad (5.129)$$

In this notation, the subscripts uu indicate unweighted averages at the first and second stages of sampling. This estimator has variance given by

$$\text{Var}(\bar{y}_{uu}) = \frac{\sigma_\alpha^2}{a} + \frac{\sigma_\beta^2}{a^2} \sum_{i=1}^a \frac{1}{b_i} + \frac{\sigma^2}{a^2} \sum_{i=1}^a \frac{1}{b_i^2} \sum_{j=1}^{b_i} \frac{1}{n_{ij}} \quad (5.130)$$

Weighting scheme 2 in Table 5.5 leads to the estimator given by

$$\bar{y}_{uw} = \frac{1}{a} \sum_{i=1}^a (y_{i.}/n_{i.}) \quad (5.131)$$

If this estimator is used in Example 5.13, the sample mean for the i^{th} lab is the weighted average $y_{i..}/n_{i.}$ of the aliquot averages, and $\bar{y}_{..w}$ is the unweighted average of the lab averages. This estimator has variance given by

$$\text{Var}(\bar{y}_{..w}) = \frac{\sigma_{\alpha}^2}{a} + \frac{\sigma_{\beta}^2}{a^2} \sum_{i=1}^a \frac{1}{n_{i.}^2} \sum_{j=1}^{b_i} n_{ij}^2 + \frac{\sigma^2}{a^2} \sum_{i=1}^a \frac{1}{n_{i.}} \quad (5.132)$$

Weighting scheme 3 in Table 5.5 leads to the estimator given by

$$\bar{y}_{wu} = \frac{1}{b.} \sum_{i=1}^a \sum_{j=1}^{b_i} (y_{ij}/n_{ij}) \quad (5.133)$$

In Example 5.13, this estimator would assign equal weight to the aliquots in computing the i^{th} lab average, but would weight the i^{th} lab average according to the number of aliquots analyzed by lab i . This estimator has variance given by

$$\text{Var}(\bar{y}_{wu}) = \frac{\sigma_{\alpha}^2}{b.^2} \sum_{i=1}^a b_i^2 + \frac{\sigma_{\beta}^2}{b.} + \frac{\sigma^2}{b.^2} \sum_{i=1}^a \sum_{j=1}^{b_i} \frac{1}{n_{ij}} \quad (5.134)$$

Finally, weighting scheme 4 in Table 5.5 leads to the estimator

$$\bar{y}_{ww} = y_{...}/n_{..} \quad (5.135)$$

which in Example 5.13 would weight the ij^{th} aliquot average according to the number of analyses in computing the i^{th} lab average and would weight the i^{th} lab average according to the number of aliquots analyzed by lab i . This estimator has variance given by

$$\text{Var}(\bar{y}_{ww}) = \frac{\sigma_{\alpha}^2}{n_{..}^2} \sum_{i=1}^a n_{i.}^2 + \frac{\sigma_{\beta}^2}{n_{..}^2} \sum_{i=1}^a \sum_{j=1}^{b_i} n_{ij}^2 + \frac{\sigma^2}{n_{..}} \quad (5.136)$$

All four of the above estimators for μ are unbiased. As noted previously, however, their variances will differ for specified values of the variance components, the b_i and the n_{ij} . The following guidelines are useful in deciding which estimator is most appropriate (has the smallest variance) for a particular application:

1. \bar{y}_{uu} is recommended when $\hat{\sigma}_\alpha^2 \geq \hat{\sigma}_\beta^2 \geq \hat{\sigma}^2$.
2. \bar{y}_{uw} is recommended when $\hat{\sigma}_\alpha^2 \geq \hat{\sigma}^2$, and $\hat{\sigma}_\beta^2$ is small relative to $\hat{\sigma}^2$.
3. \bar{y}_{wu} is recommended when $\hat{\sigma}_\beta^2 \geq \hat{\sigma}^2$, and $\hat{\sigma}_\alpha^2$ is small relative to $\hat{\sigma}_\beta^2$.
4. \bar{y}_{ww} is recommended when $\hat{\sigma}_\alpha^2 < \hat{\sigma}_\beta^2 < \hat{\sigma}^2$, where $\hat{\sigma}_\alpha^2$ is small relative to $\hat{\sigma}_\beta^2$, and $\hat{\sigma}_\beta^2$ is small relative to $\hat{\sigma}^2$.

When a case is encountered that does not satisfy the criteria for any of the above guidelines, the estimated variances of the four estimators should be evaluated (formulas are given below), and the estimator with the smallest estimated variance should be used.

When the unequal b_i and n_{ij} are planned in advance for the specific purpose of utilizing one of the weighting schemes in Table 5.5, the corresponding estimator of μ will be unbiased but might have a larger variance than one or more of the other estimators. If this is found to be true, the investigator should reconsider the motivation and objectives that led to the choice of a particular weighting scheme. If the difference in variances is large, then perhaps obtaining a more precise estimate of μ (one with a smaller variance) is more desirable than achieving a particular weighting of the observations in estimating μ .

The variances given by Equations 5.130, 5.132, 5.134, and 5.136 are estimated by replacing the variance components σ^2 , σ_β^2 , and σ_α^2 with their respective estimates $\hat{\sigma}^2$, $\hat{\sigma}_\beta^2$, $\hat{\sigma}_\alpha^2$ computed from Equations 5.118, 5.119, and 5.120. This yields a linear combination of the variance component estimates of the form

$$\text{Vâr}(\text{Estimator of } \mu) = a_1 \hat{\sigma}_\alpha^2 + a_2 \hat{\sigma}_\beta^2 + a_3 \hat{\sigma}^2 \quad (5.137)$$

Equation 5.137 can be expressed as a linear combination of the computed mean squares MS_α , MS_β , and MS_ϵ by replacing $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_\beta^2$, and $\hat{\sigma}^2$ with the terms on the right-hand sides of Equations 5.118, 5.119, and 5.120. This gives

$$\begin{aligned} \text{Vâr}(\text{Estimator of } \mu) &= \frac{a_1}{k_2} MS_\alpha - \frac{a_1 k_1}{k_2 k_3} MS_\beta + a_1 \left(\frac{k_1 - k_3}{k_2 k_3} \right) MS_\epsilon \\ &\quad + \frac{a_2}{k_3} MS_\beta - \frac{a_2}{k_3} MS_\epsilon + a_3 MS_\epsilon \\ &= \frac{a_1}{k_2} MS_\alpha + \left(\frac{a_2}{k_3} - \frac{a_1 k_1}{k_2 k_3} \right) MS_\beta \end{aligned}$$

$$\begin{aligned}
 & + \left[a_1 \left(\frac{k_1 - k_3}{k_2 k_3} \right) - \frac{a_2}{k_3} + a_3 \right] MS_\epsilon \\
 & = m_3 MS_\alpha + m_4 MS_\beta + m_5 MS_\epsilon \quad (5.138)
 \end{aligned}$$

Using Satterthwaite's (1946) formula, the approximate degrees of freedom for an estimator of this form are given by

$$\hat{df} = \frac{(m_3 MS_\alpha + m_4 MS_\beta + m_5 MS_\epsilon)^2}{\frac{(m_3 MS_\alpha)^2}{df_\alpha} + \frac{(m_4 MS_\beta)^2}{df_\beta} + \frac{(m_5 MS_\epsilon)^2}{df_\epsilon}} \quad (5.139)$$

Application of the concepts and some of the equations presented in this section are illustrated in Example 5.14.

If it is assumed that y_{ijk} is normally distributed with mean μ and variance $\sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$, then each of the estimators of μ given by Equations 5.129, 5.131, 5.133, and 5.135 is normally distributed with mean μ and with variance as specified for each estimator. Thus, whichever of the four estimators of μ is chosen for a given application, an approximate $100(1-\alpha)\%$ confidence interval for μ is given by

$$(\text{Estimate of } \mu) \pm \sqrt{\hat{V}\hat{a}r(\text{Estimator of } \mu)} t_{1-\alpha/2}(\hat{df}) \quad (5.140)$$

where \hat{df} is given by Equation 5.139 and $t_{1-\alpha/2}(\hat{df})$ is obtained from Table A5.

Also, hypotheses about μ are tested by using methods analogous to those in Section 4.2.2. The test statistic is given by

$$t = \frac{(\text{Estimate of } \mu) - \mu_0}{\sqrt{\hat{V}\hat{a}r(\text{Estimator of } \mu)}} \quad (5.141)$$

In this formulation, μ_0 is a hypothesized value of μ , and t is compared with a value obtained from Table A5 with degrees of freedom \hat{df} .

Example 5.14 For the experiment described in Example 5.13, construct a 95% confidence interval for the overall mean μ . From the results of Example 5.13, note that $\hat{\sigma}_{\text{Labs}}^2 > \hat{\sigma}_{\text{Aliq}}^2 > \hat{\sigma}_{\text{Anal}}^2$; thus, the unweighted estimator \bar{y}_{uu} , given by Equation 5.129, is appropriate.

Substituting the data values from Example 5.13 into Equation 5.129 yields

$$\begin{aligned}\bar{y}_{uu} &= \frac{1}{3} \left[\frac{1}{3} \left(\frac{0.722}{4} + \frac{1.014}{3} + \frac{1.124}{3} \right) \right. \\ &\quad \left. + \frac{1}{3} \left(\frac{0.847}{1} + \frac{1.844}{2} + \frac{1.494}{2} \right) \right. \\ &\quad \left. + \frac{1}{2} \left(\frac{1.853}{3} + \frac{1.222}{2} \right) \right] = 0.5836\end{aligned}$$

To estimate the variance of \bar{y}_{uu} by applying Equation 5.130, the quantities

$$\frac{1}{a^2} \sum_{i=1}^a \frac{1}{b_i} = \frac{1}{3^2} \left(\frac{1}{3} + \frac{1}{3} + \frac{1}{2} \right) = \frac{7}{54} = 0.129630$$

and

$$\begin{aligned}\frac{1}{a^2} \sum_{i=1}^a \frac{1}{b_i^2} \sum_{j=1}^{b_i} \frac{1}{n_{ij}} &= \frac{1}{3^2} \left[\frac{1}{3^2} \left(\frac{1}{4} + \frac{1}{3} + \frac{1}{3} \right) \right. \\ &\quad \left. + \frac{1}{3^2} \left(\frac{1}{1} + \frac{1}{2} + \frac{1}{2} \right) + \frac{1}{2^2} \left(\frac{1}{3} + \frac{1}{2} \right) \right] \\ &= \frac{1}{9} \left(\frac{11}{108} + \frac{2}{9} + \frac{5}{24} \right) = 0.059156\end{aligned}$$

are computed from the Example 5.13 data. The variance component estimates computed in Example 5.13 are

$$\hat{\sigma}_{\text{Labs}}^2 = 0.083801$$

$$\hat{\sigma}_{\text{Aliq}}^2 = 0.007411$$

and

$$\hat{\sigma}_{\text{Anal}}^2 = 0.003322$$

By substituting these values into Equation 5.130, the variance estimate is

$$\begin{aligned}\text{Vâr}(\bar{y}_{uu}) &= \frac{0.083801}{3} + (0.129630)(0.007411) \\ &\quad + (0.059156)(0.003322) \\ &= 0.029091\end{aligned}$$

Note that this is Equation 5.137 with $a_1 = 1/3$, $a_2 = 0.129630$, and $a_3 = 0.059156$.

Recall that the values $k_1 = 2.50$, $k_2 = 6.25$, and $k_3 = 2.44$ were computed in Example 5.13. Substituting these values into the individual terms of Equation 5.138 yields

$$\begin{aligned}m_3 &= \frac{a_1}{k_2} = \frac{1}{3(6.25)} = 0.053333 \\ m_4 &= \frac{7}{54(2.44)} - \frac{2.50}{3(6.25)(2.44)} = -0.001518\end{aligned}$$

and

$$m_5 = \frac{1}{3} \left(\frac{2.50 - 2.44}{6.25(2.44)} \right) - \frac{7}{54(2.44)} + 0.059156 = 0.007341$$

Thus, in terms of the computed mean squares, the variance estimate is

$$\begin{aligned}\text{Vâr}(\bar{y}_{uu}) &= 0.053333 \text{ MS}_{\text{Labs}} - 0.001518 \text{ MS}_{\text{Aliq}} + 0.007341 \text{ MS}_{\text{Anal}} \\ &= (0.053333)(0.545605) - (0.001518)(0.021404) \\ &\quad + (0.007341)(0.003322) = 0.029091\end{aligned}$$

This is, of course, the same value as calculated above.

The degrees of freedom for $\text{Vâr}(\bar{y}_{uu})$ are obtained by substituting m_3 , m_4 , m_5 , the computed mean squares, and their degrees of freedom from Example 5.13 into Equation 5.139. By doing so, the numerator is

$$\begin{aligned}&[0.053333(0.545605) - 0.001518(0.021404) \\ &\quad + 0.007341(0.003322)]^2 = 0.0008436\end{aligned}$$

and the denominator is

$$\frac{[0.053333(0.545605)]^2}{2} + \frac{[-0.001518(0.021404)]^2}{5} + \frac{[0.007341(0.003322)]^2}{12} = 0.0004234$$

Thus, application of Equation 5.139 yields

$$\hat{df} = \frac{0.0008436}{0.0004234} = 1.99 \cong 2$$

From Table A5, the value $t_{0.975}(2) = 4.30$ is obtained, and by applying Equation 5.140, a 95% confidence interval for μ is given by

$$0.5836 \pm \sqrt{0.029091} (4.30)$$

which yields the interval $(-0.1498, 1.3170)$. Note that the data are in percent recovery minus 86, so that a negative value is meaningful. In terms of the original percent recovery units, a 95% confidence interval for the unknown true mean is $(85.8502, 87.3170)$.

5.3.1.3 Equal Sample Sizes at Each Stage

When all n_{ij} are equal, say $n_{ij} = n$, and all b_i are equal, say $b_i = b$, the methods presented in Sections 5.3.1.1 and 5.3.1.2 are considerably simplified. Equations 5.104 through 5.117 can be simplified, and the resulting analysis of variance table is shown in Table 5.6.

TABLE 5.6
Nested Random Effects Analysis of
Variance Table with $n_{ij} = n$ and $b_i = b$

Source	df	SS	MS	E[MS]	F
α_i	$a-1$	SS_α	MS_α	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\alpha^2$	MS_α/MS_β
$\beta_{(i)j}$	$a(b-1)$	SS_β	MS_β	$\sigma^2 + n\sigma_\beta^2$	MS_β/MS_ϵ
$\epsilon_{(ij)k}$	$ab(n-1)$	SS_ϵ	MS_ϵ	σ^2	

Note from Table 5.6 that the coefficients for σ_β^2 are the same in both $E[MS_\alpha]$ and $E[MS_\beta]$; i.e., $k_1 = k_3 = n$. Thus, an exact test of the null hypothesis $H_0: \sigma_\alpha^2 = 0$ can be performed by computing the ratio

$$F = MS_\alpha/MS_\beta$$

The null hypothesis is rejected at the α level of significance if F is greater than or equal to $F_{1-\alpha}(a-1, ab-a)$, which is obtained from Table A6. The null hypothesis $H_0: \sigma_{\beta}^2 = 0$ is tested exactly as described in Section 5.3.1.1.

Estimates of the variance components are computed by applying Equations 5.118, 5.119, 5.120. Note, however, that $k_1 = k_3 = n$ and $k_2 = bn$.

When $n_{ij} = n$ and $b_i = b$, all four estimators of μ , presented in Section 5.3.1.2, are equivalent to the estimator

$$\bar{y}_{...} = y_{...}/abn \quad (5.142)$$

which has variance given by

$$\text{Var}(\bar{y}_{...}) = \frac{\sigma^2 + n\sigma_{\beta}^2 + bn\sigma_{\alpha}^2}{abn} = \frac{E[MS_{\alpha}]}{abn} \quad (5.143)$$

Thus, $\text{Var}(\bar{y}_{...})$ is estimated by

$$\hat{\text{Var}}(\bar{y}_{...}) = \frac{MS_{\alpha}}{abn} \quad (5.144)$$

which has $a - 1$ degrees of freedom.

5.3.2 The Nested Mixed Effects Model

Suppose now that in the example discussed in introducing Equation 5.101, the four batches of pellets are *not* assumed to be a random sample from all batches of pellets produced by the process. Instead, suppose that the four batches of pellets were manufactured under four different sets of operating conditions, possibly representing changes in the process over a given period, and it is of interest to estimate and compare the mean percent plutonium values for the four batches. It is still assumed that a random sample of six pellets is taken from each batch and that the three analyses performed on a given pellet are a random sample from the population of all possible analyses of the pellet. The appropriate model is given by Equation 5.101, where the terms in the model are defined as follows:

$\mu + \alpha_i = \mu_i$ = the mean percent plutonium for the population of pellets in the i^{th} batch. That is, the α_i are fixed effects.

$\beta_{(ij)}$ = the deviation from μ_i due to the ij^{th} pellet; $\beta_{(ij)}$ is a random observation from a population with zero mean and variance σ_{β}^2 .

$\epsilon_{(ij)k}$ = the deviation from $\mu_i + \beta_{(i)j}$ due to the k^{th} analysis on the ij^{th} pellet; $\epsilon_{(ij)k}$ is a random observation from a population with zero mean and with variance σ^2 .

It is also assumed that the $\beta_{(i)j}$ are uncorrelated, the $\epsilon_{(ij)k}$ are uncorrelated, and there are no correlations among the $\beta_{(i)j}$ and the $\epsilon_{(ij)k}$.

Thus, y_{ijk} is an observation from a population with mean $\mu_i = \mu + \alpha_i$ and with variance $\sigma_{\beta}^2 + \sigma^2$. In general, with the nested mixed effects model, it is of interest to

1. Estimate and compare the means μ_i .
2. Estimate σ_{β}^2 and σ^2 .
3. Test the hypothesis $H_0: \sigma_{\beta}^2 = 0$.

5.3.2.1 An Overall Analysis of Variance

As in Section 5.3.1.1, all quantities given by Equations 5.104 through 5.117, except Equation 5.116, are computed. An analysis of variance table is then constructed as shown in Table 5.7. Note that $E[MS_{\alpha}]$ is different in Table 5.7 than in Table 5.4. In Table 5.7, the quantity $\bar{\alpha}$ in $E[MS_{\alpha}]$ is defined as

$$\bar{\alpha} = \sum_{i=1}^a n_i \alpha_i / n_{..} \tag{5.145}$$

where n_i and $n_{..}$ are given by Equations 5.105.

TABLE 5.7
Nested Mixed Effects Analysis of Variance Table

Source	df	SS	MS	E[MS]
α_i	$a - 1$	SS_{α}	MS_{α}	$\sigma^2 + k_1 \sigma_{\beta}^2 + \frac{1}{a-1} \sum_{i=1}^a n_i (\alpha_i - \bar{\alpha})^2$
$\beta_{(i)j}$	$b - a$	SS_{β}	MS_{β}	$\sigma^2 + k_3 \sigma_{\beta}^2$
$\epsilon_{(ij)k}$	$n_{..} - b$	SS_{ϵ}	MS_{ϵ}	σ^2

Consider now testing the hypothesis

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

against the alternative

$$\tag{5.146}$$

H_A : some of the means are not equal

These hypotheses can be rewritten in the equivalent form

$$\begin{aligned} H_0: \alpha_i &= 0, \quad i = 1, 2, \dots, a \\ H_A: \alpha_i &\neq 0, \quad \text{for some } i \end{aligned} \quad (5.147)$$

The procedure for testing the hypotheses in Equation 5.146 or 5.147 is identical to the approximate procedure presented in connection with Equations 5.123 through 5.126 for testing the hypothesis that $\sigma_\alpha^2 = 0$ in the nested random effects model.

The hypothesis that $\sigma_\beta^2 = 0$ is tested by the procedure presented in connection with Equations 5.121 and 5.122. Estimates of the variance components σ^2 and σ_β^2 are computed by applying Equations 5.118 and 5.119.

5.3.2.2 Estimation and Comparison of the Means $\mu_i = \mu + \alpha_i$

When the hypothesis of equal means given by Equation 5.146 or 5.147 is not rejected, the problem can often be reformulated as one where the a classes are treated as having been randomly selected from a single population with an overall mean μ . Then the appropriate model is the nested random effects model described in Section 5.3.1, where σ_α^2 is interpreted as the variance due to sampling error and the methods described in Section 5.3.1.2 are applied to estimate the overall mean μ . If the estimate of σ_α^2 is near zero, an alternate approach is to drop α_i from the model and treat the b subclasses as a random sample of classes from a population with an overall mean μ . Then the one-way random effects analysis of variance methods described in Section 5.2 would be used to estimate the overall mean μ and the two variance components σ_β^2 and σ^2 .

When the hypothesis of equal means is rejected, however, it is of interest to estimate and compare the individual means μ_i . Methods for estimating the means μ_i from the nested mixed effects model are analogous to the methods presented in Section 5.2.3 for estimating μ from the one-way random effects model. Also, the rationale for choosing an appropriate estimator for μ_i when the n_{ij} are unequal is analogous to the rationale discussed in that section for choosing an estimator for μ . Thus, to avoid redundancy, two estimators for μ_i are presented in this section, with some brief guidelines for deciding which one to use in a particular application, but little discussion is given. The reader can review Section 5.2.3 for further discussions, keeping in mind that μ_i , σ_β^2 , and n_{ij} from this section correspond to μ , σ_α^2 , and n_i in Section 5.2.3. This section also presents methods for constructing confidence intervals and for making comparisons among the μ_i . These methods are analogous to those in Section 5.1.4. When the n_{ij} and b_i are not constant over the experiment, however, estimation of the μ_i and

computation of estimated variances and approximate degrees of freedom are somewhat more complicated for the nested mixed effects model than for the one-way fixed effects model.

If the n_{ij} are unequal and $\hat{\sigma}_\beta^2 \geq \hat{\sigma}^2$, then the averages \bar{y}_{ij} at the second stage of sampling should be weighted equally in estimating μ_i , and

$$\bar{y}_{iu} = \sum_{j=1}^{b_i} \bar{y}_{ij} / b_i = \frac{1}{b_i} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} y_{ijk} / n_{ij} \quad (5.148)$$

is the recommended estimator. The variance of \bar{y}_{iu} is given by

$$\text{Var}(\bar{y}_{iu}) = \frac{\sigma_\beta^2 + r_{ii} \sigma^2}{b_i} \quad (5.149)$$

where

$$r_{ii} = \frac{1}{b_i} \sum_{j=1}^{b_i} 1/n_{ij} \quad (5.150)$$

$\text{Var}(\bar{y}_{iu})$ is estimated by substituting the estimates $\hat{\sigma}^2$ and $\hat{\sigma}_\beta^2$, computed from Equations 5.118 and 5.119, in place of σ^2 and σ_β^2 in Equation 5.149; that is,

$$\text{V}\hat{\text{ar}}(\bar{y}_{iu}) = \frac{\hat{\sigma}_\beta^2 + r_{ii} \hat{\sigma}^2}{b_i} \quad (5.151)$$

To obtain the approximate degrees of freedom for $\text{V}\hat{\text{ar}}(\bar{y}_{iu})$, Equation 5.151 is expressed as a linear combination of MS_ϵ and MS_β . This is achieved by substituting Equations 5.118 and 5.119 into Equation 5.151, which gives

$$\begin{aligned} \text{V}\hat{\text{ar}}(\bar{y}_{iu}) &= \frac{MS_\beta - MS_\epsilon}{b_i k_3} + \frac{r_{ii} MS_\epsilon}{b_i} \\ &= m_{i1} MS_\beta + m_{i2} MS_\epsilon \end{aligned} \quad (5.152)$$

where

$$m_{i1} = \frac{1}{b_i k_3} \quad (5.153)$$

and

$$m_{i2} = \frac{(r_{i1} k_3 - 1)}{b_i k_3} \quad (5.154)$$

Then, by using Satterthwaite's (1946) formula, the approximate degrees of freedom for $\hat{V}\hat{a}r(\bar{y}_{iu})$ are given by

$$\hat{d}f_{iu} = \frac{(m_{i1} MS_\beta + m_{i2} MS_\epsilon)^2}{\frac{(m_{i1} MS_\beta)^2}{df_\beta} + \frac{(m_{i2} MS_\epsilon)^2}{df_\epsilon}} \quad (5.155)$$

where $\hat{d}f_{iu}$ is rounded to the nearest integer. When it is desirable to be conservative, $\hat{d}f_{iu}$ is rounded down.

When $\hat{\sigma}_\beta^2$ is small relative to $\hat{\sigma}^2$, the weighted average

$$\bar{y}_{iw} = y_{i.}/n_i = \frac{1}{n_i} \sum_{j=1}^{b_i} n_{ij} \bar{y}_{ij} \quad (5.156)$$

is the recommended estimator. The variance of \bar{y}_{iw} is given by

$$\text{Var}(\bar{y}_{iw}) = \frac{r_{i2} \sigma_\beta^2 + \sigma^2}{n_i} \quad (5.157)$$

where

$$r_{i2} = \sum_{j=1}^{b_i} n_{ij}^2 / n_i \quad (5.158)$$

$\text{Var}(\bar{y}_{iw})$ is estimated by substituting the estimates $\hat{\sigma}^2$ and $\hat{\sigma}_\beta^2$, computed from Equations 5.118 and 5.119, in place of σ^2 and σ_β^2 in Equation 5.157; that is,

$$\hat{V}\hat{a}r(\bar{y}_{iw}) = \frac{r_{i2} \hat{\sigma}_\beta^2 + \hat{\sigma}^2}{n_i} \quad (5.159)$$

To obtain the approximate degrees of freedom for $\hat{V}\hat{a}r(\bar{y}_{iw})$, note that substituting Equations 5.118 and 5.119 into Equation 5.159 gives

$$\begin{aligned} \hat{V}\hat{a}r(\bar{y}_{iw}) &= \frac{r_{i2}(MS_\beta - MS_\epsilon)}{n_i k_3} + \frac{MS_\epsilon}{n_i} \\ &= m_{i3} MS_\beta + m_{i4} MS_\epsilon \end{aligned} \quad (5.160)$$

where

$$m_{i3} = \frac{r_{i2}}{n_i k_3} \quad (5.161)$$

and

$$m_{i4} = \frac{1}{n_i} \left(1 - \frac{r_{i2}}{k_3} \right) \quad (5.162)$$

The approximate degrees of freedom for $\hat{V}\hat{a}r(\bar{y}_{iw})$ are then given by

$$\hat{d}f_{iw} = \frac{(m_{i3} MS_\beta + m_{i4} MS_\epsilon)^2}{\frac{(m_{i3} MS_\beta)^2}{df_\beta} + \frac{(m_{i4} MS_\epsilon)^2}{df_\epsilon}} \quad (5.163)$$

where $\hat{d}f_{iw}$ is rounded to the nearest integer. When it is desirable to be conservative, $\hat{d}f_{iw}$ is rounded down.

In this discussion, \bar{y}_{iu} is recommended when $\hat{\sigma}_\beta^2 \geq \hat{\sigma}^2$, and \bar{y}_{iw} is recommended when $\hat{\sigma}_\beta^2$ is small relative to $\hat{\sigma}^2$. When cases are encountered that do not satisfy either of the above criteria, the estimated variances of \bar{y}_{iu} and \bar{y}_{iw} should be computed for all i . Then, for each i , the estimator with the smallest estimated variance should be used.

If it is assumed that y_{ijk} is an observation from a $N(\mu_i, \sigma_\beta^2 + \sigma^2)$ distribution, then each of the estimators of μ_i given by Equations 5.148 and 5.156 is normally distributed with mean μ_i and with variance as specified for each estimator. Thus, whichever estimator is chosen for a given application, an approximate $100(1 - \alpha)\%$ confidence interval for μ_i is given by

$$(\text{Estimate of } \mu_i) \pm \sqrt{\hat{V}\hat{a}r(\text{Estimator of } \mu_i)} t_{1-\alpha/2}(\hat{d}f_i) \quad (5.164)$$

where $\hat{d}f_i$ is computed from Equation 5.155 or 5.163, depending upon which estimator of μ_i is used, and $t_{1-\alpha/2}(\hat{d}f_i)$ is obtained from Table A5.

The problem of constructing simultaneous confidence intervals for the a means μ_i is dealt with by using one of the two methods introduced in Section 5.1.4.1, with some slight modifications. By using Scheffé's (1959) method, the formula

$$(\text{Estimate of } \mu_i) \pm \sqrt{(a-1)F_{1-\alpha}(a-1, \hat{d}f_i) \hat{V}\hat{a}r(\text{Estimator of } \mu_i)} \quad (5.165)$$

yields confidence intervals for the a class means with an overall probability

of at least $(1-\alpha)$ that the a intervals simultaneously cover their respective true means. In this formulation, (Estimate of μ_i), $\text{V}\hat{\text{a}}\text{r}(\text{Estimator of } \mu_i)$, and $\hat{\text{d}}\hat{\text{f}}_i$ are computed from Equations 5.148, 5.151, and 5.155 if \bar{y}_{iu} is the chosen estimator of μ_i , or from Equations 5.156, 5.159, and 5.163 if \bar{y}_{iw} is chosen. Also, $F_{1-\alpha}(a-1, \hat{\text{d}}\hat{\text{f}}_i)$ is obtained from Table A6.

The Bonferroni method guarantees that using α/a in place of α in Equation 5.164 will yield confidence intervals for the a means with an overall probability of at least $(1-\alpha)$ that the a intervals simultaneously cover their respective true means. The appropriate formula is

$$(\text{Estimate of } \mu_i) \pm t_{(1-\alpha/2a)}(\hat{\text{d}}\hat{\text{f}}_i) \sqrt{\text{V}\hat{\text{a}}\text{r}(\text{Estimator of } \mu_i)} \quad (5.166)$$

Either method guarantees an overall confidence level of at least $(1-\alpha)$. The width of the resulting intervals, however, will not in general be the same for both methods. Thus, it may be desirable to construct intervals by both methods and then report the narrower intervals.

In most applications, comparisons of the means μ_i can be expressed as contrasts of the form

$$C = \sum_{i=1}^a C_i \mu_i \quad (5.167)$$

where

$$\sum_{i=1}^a C_i = 0 \quad (5.168)$$

Note also that Equation 5.167 can be expressed as

$$\begin{aligned} C &= \sum_{i=1}^a C_i \mu_i = \sum_{i=1}^a C_i (\mu + \alpha_i) \\ &= \mu \left(\sum_{i=1}^a C_i \right) + \sum_{i=1}^a C_i \alpha_i \\ &= \sum_{i=1}^a C_i \alpha_i \end{aligned} \quad (5.169)$$

Whether working in terms of Equation 5.167 or 5.169, an estimate of C is obtained by substituting either \bar{y}_{iu} or \bar{y}_{iw} for μ_i ; that is,

$$\hat{C} = \sum_{i=1}^a C_i (\text{Estimate of } \mu_i) \quad (5.170)$$

The variance of \hat{C} is

$$\text{Var}(\hat{C}) = \sum_{i=1}^a C_i^2 \text{Var}(\text{Estimator of } \mu_i) \quad (5.171)$$

which is estimated by substituting either $\text{V}\hat{\text{ar}}(\bar{y}_{iu})$ or $\text{V}\hat{\text{ar}}(\bar{y}_{iw})$ for $\text{Var}(\text{Estimator of } \mu_i)$. That is,

$$\text{V}\hat{\text{ar}}(\hat{C}) = \sum_{i=1}^a C_i^2 \text{V}\hat{\text{ar}}(\text{Estimator of } \mu_i) \quad (5.172)$$

To accommodate computing approximate degrees of freedom for $\text{V}\hat{\text{ar}}(\hat{C})$, either Equation 5.152 or 5.160 is substituted for $\text{V}\hat{\text{ar}}(\text{Estimator of } \mu_i)$ in Equation 5.172. Thus, if \bar{y}_{iu} is chosen to estimate μ_i , then

$$\text{V}\hat{\text{ar}}(\hat{C}) = \left[\sum_{i=1}^a C_i^2 m_{i1} \right] \text{MS}_\beta + \left[\sum_{i=1}^a C_i^2 m_{i2} \right] \text{MS}_\epsilon \quad (5.173)$$

If \bar{y}_{iw} is chosen to estimate μ_i , then

$$\text{V}\hat{\text{ar}}(\hat{C}) = \left[\sum_{i=1}^a C_i^2 m_{i3} \right] \text{MS}_\beta + \left[\sum_{i=1}^a C_i^2 m_{i4} \right] \text{MS}_\epsilon \quad (5.174)$$

In either case, $\text{V}\hat{\text{ar}}(\hat{C})$ is a linear combination of the computed mean squares, MS_β and MS_ϵ , which has the form

$$\text{V}\hat{\text{ar}}(\hat{C}) = a_1 \text{MS}_\beta + a_2 \text{MS}_\epsilon \quad (5.175)$$

Thus, the approximate degrees of freedom for $\text{V}\hat{\text{ar}}(\hat{C})$, computed by Satterthwaite's (1946) method, are

$$\hat{df} = \frac{(a_1 \text{MS}_\beta + a_2 \text{MS}_\epsilon)^2}{\frac{(a_1 \text{MS}_\beta)^2}{df_\beta} + \frac{(a_2 \text{MS}_\epsilon)^2}{df_\epsilon}} \quad (5.176)$$

An approximate $100(1-\alpha)\%$ confidence interval for C is given by

$$\hat{C} \pm t_{1-\alpha/2}(\hat{df}) \sqrt{\hat{\text{Var}}(\hat{C})} \quad (5.177)$$

where $t_{1-\alpha/2}(\hat{df})$ is obtained from Table A5. To test the hypothesis

$$H_0: \sum_{i=1}^a C_i \mu_i = 0$$

against the alternative (5.178)

$$H_A: \sum_{i=1}^a C_i \mu_i = C_A \neq 0$$

compute the test statistic

$$t = \frac{\hat{C}}{\sqrt{\hat{\text{Var}}(\hat{C})}} \quad (5.179)$$

H_0 is rejected at the α level of significance if

$$t \geq t_{1-\alpha/2}(\hat{df})$$

or (5.180)

$$t \leq -t_{1-\alpha/2}(\hat{df})$$

where $t_{1-\alpha/2}(\hat{df})$ is obtained from Table A5.

These confidence interval and hypothesis testing methods are recommended for making any comparisons among the means that were planned before collection of the data; that is, comparisons the experiment was specifically designed to make. Sometimes when the estimated means are examined, patterns become apparent which suggest certain comparisons that were not planned. The corresponding contrasts are estimated by applying Equation 5.170, with variance estimates and approximate degrees of freedom computed from Equations 5.172 and 5.176. The confidence interval and hypothesis test described above are not valid, however, because the contrasts were selected after examining the data. In such

cases, Scheffé's (1959) method is recommended. Any number of contrast hypotheses can be simultaneously tested by using Scheffé's method, and the overall probability of making a Type I error is at most α . Similarly, confidence intervals can be constructed for any number of contrasts, with an overall probability of at least $(1 - \alpha)$ that the intervals simultaneously cover the corresponding contrasts of the true means μ_i .

Scheffé confidence intervals for any number of contrasts are constructed from the formula

$$\hat{C} \pm \sqrt{(a-1)F_{1-\alpha}(a-1, df) \text{Vâr}(\hat{C})} \tag{5.181}$$

where \hat{C} , $\text{Vâr}(\hat{C})$, and df are computed from Equations 5.170, 5.172, and 5.176, and where $F_{1-\alpha}(a-1, df)$ is obtained from Table A6.

To test hypotheses of the form given by Equation 5.178 for any number of contrasts, the test statistic t is computed for each contrast by applying Equation 5.179. For a given contrast, H_0 is rejected if

$$t \geq \sqrt{(a-1)F_{1-\alpha}(a-1, df)}$$

or

$$\tag{5.182}$$

$$t \leq -\sqrt{(a-1)F_{1-\alpha}(a-1, df)}$$

5.3.2.3 Equal Sample Sizes at Each Stage

When all n_{ij} are equal ($n_{ij} = n$) and all b_i are equal ($b_i = b$) the methods presented in Sections 5.3.2.1 and 5.3.2.2 are considerably simplified. Equations 5.104 through 5.117 simplify, and the form of the resulting analysis of variance table is illustrated by Table 5.8.

TABLE 5.8
Nested Mixed Effects Analysis of Variance
Table with $n_{ij} = n$ and $b_i = b$

Source	df	SS	MS	E[MS]	F
α_i	$a-1$	SS_α	MS_α	$\sigma^2 + n\sigma_\beta^2 + \frac{bn}{a-1} \sum_{i=1}^a \alpha_i^2$	MS_α/MS_β
$\beta_{(ij)}$	$a(b-1)$	SS_β	MS_β	$\sigma^2 + n\sigma_\beta^2$	MS_β/MS_ϵ
$\epsilon_{(ijk)}$	$ab(n-1)$	SS_ϵ	MS_ϵ	σ^2	

Note from Table 5.8 that the coefficients for σ_β^2 are the same in both $E[MS_\alpha]$ and $E[MS_\beta]$; that is, $k_1 = k_3 = n$. Thus, an exact test of the null hypothesis given by Equation 5.146 or 5.147 can be performed by computing the ratio

$$F = MS_\alpha / MS_\beta$$

The null hypothesis is rejected at the α level of significance if F is greater than or equal to $F_{1-\alpha}(a-1, ab-a)$, which is obtained from Table A.6. The null hypothesis $H_0: \sigma_\beta^2 = 0$ is tested exactly as described in Section 5.3.1.1. Estimates of the variance components σ^2 and σ_β^2 are computed by applying Equations 5.118 and 5.119, where $k_3 = n$.

When $n_{ij} = n$ and $b_i = b$, both estimators of μ_i , given by Equations 5.148 and 5.156, reduce to the estimator

$$\bar{y}_{i..} = y_{i..} / bn = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk} \quad (5.183)$$

which has variance given by

$$\text{Var}(\bar{y}_{i..}) = \frac{\sigma^2 + n\sigma_\beta^2}{bn} = \frac{E[MS_\beta]}{bn} \quad (5.184)$$

Thus, $\text{Var}(\bar{y}_{i..})$ is estimated by

$$\hat{\text{Var}}(\bar{y}_{i..}) = \frac{MS_\beta}{bn} \quad (5.185)$$

which has $a(b-1)$ degrees of freedom.

When computing confidence intervals for individual μ_i or for contrasts, the degrees of freedom, $\hat{d}f_i$ or $\hat{d}f$, in Equations 5.164, 5.165, 5.166, 5.177, and 5.181 are replaced by $a(b-1)$. The same is true for the hypothesis testing criteria given by Equations 5.180 and 5.182.

Duncan's multiple range test, presented in Section 5.1.4.3, can be readily applied for comparing all possible pairs of the sample means $\bar{y}_{i..}$ when the n_{ij} and b_i are constant over the experiment. The only modifications to the method presented in Section 5.1.4.3 are that $\bar{y}_{i..}$ is used in place of \bar{y}_i , MS_β/bn is used in place of MS_{within}/n , and $a(b-1)$ is used in place of $N-a$. Note, however, that the approximate procedure described by Equations 5.58 through 5.60 does not apply directly. In such cases, it is recommended that Scheffé's procedure be applied to test for differences between all possible pairs of means.

5.4 ANALYSIS OF VARIANCE WITH TWO-WAY CROSSED CLASSIFICATIONS

The conceptual definition of a two-way crossed classification experimental design is that two factors are studied by including all combinations of two or more levels of each factor in an experiment. Then a random sample of observed or measured values of a response variable is taken at each combination of the factor levels. Experimental data obtained in this manner provide information for assessing the effects of the levels of each factor on the response variable and for assessing how the two factors interact to affect the response variable. The reader is referred to Chapter 7 for guidance on designing and conducting an experiment with two factors.

The term "crossed classification" is used to distinguish the factorial type of experimental designs, such as those described in Section 7.3, from the nested or hierarchical type of experimental designs, such as those described in Sections 5.3 and 7.6. Some of the basic concepts underlying the two-way crossed classification are introduced and discussed in Section 7.3.1. The randomized complete block design introduced in Section 7.2.7, however, is also a two-way crossed classification, as are some other special experimental designs.

This section provides computational formulas for performing an analysis of variance using data from a general two-way crossed classification design with equal sample sizes at all combinations of the factor levels. The specific use and interpretation of the computed quantities in parameter estimation and hypothesis testing depend upon whether the effects of the factors are assumed to be fixed or random. Thus, estimation and hypothesis testing methods are presented for three sets of assumptions:

1. Both factors have fixed effects.
2. Both factors have random effects.
3. One factor has fixed effects and the other has random effects.

5.4.1 Computational Formulas

Assuming that the two factors are referred to as factor A and factor B, the data from a two-way crossed classification design can be displayed as shown in Table 5.9. The linear model that characterizes the observed or measured response values is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \quad (5.186)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, n$$

The terms in this model are defined as follows:

- μ = the overall mean
- α_i = the main effect of the i^{th} level of factor A averaged over the levels of factor B
- β_j = the main effect of the j^{th} level of factor B averaged over the levels of factor A
- $\alpha\beta_{ij}$ = the effect of the interaction of the i^{th} level of factor A with the j^{th} level of factor B
- ϵ_{ijk} = the effect of random error on the k^{th} observed or measured response taken at the i^{th} level of factor A and j^{th} level of factor B.

TABLE 5.9

Sample Data Layout for a Two-Way
Crossed Classification Design

Factor A levels	Factor B levels			
	1	2	...	b
1	Y_{111}	Y_{121}	...	Y_{1b1}
	Y_{112}	Y_{122}		Y_{1b2}
	.	.		.
	.	.		.
	Y_{11n}	Y_{12n}		Y_{1bn}
2	Y_{211}	Y_{221}		Y_{2b1}
	Y_{212}	Y_{222}		Y_{2b2}
	.	.		.
	.	.		.
	Y_{21n}	Y_{22n}		Y_{2bn}
.	.	.	.	
.	.	.	.	
.	.	.	.	
a	Y_{a11}	Y_{a21}		Y_{ab1}
	Y_{a12}	Y_{a22}		Y_{ab2}
	.	.		.
	.	.		.
	Y_{a1n}	Y_{a2n}		Y_{abn}

In this section, it is assumed that $n > 0$ observed or measured responses have been taken at each of the ab combinations of the levels of factor A and factor B. The corresponding experimental design is referred to as a **balanced, complete design**. The resulting data from such an experiment are often called **balanced data**.

In many experimental situations, it is not possible to perform a balanced, complete experiment. When this happens, the resulting data are called **unbalanced** if different numbers of observed or measured responses are taken at the ab factor level combinations. If any of the ab factor level combinations have no response values, the experimental design is said to be **incomplete**. The general analysis of variance methods for unbalanced and/or incomplete two-way designs is presented in textbooks on linear models, such as Bancroft (1968), Hocking (1985), Milliken and Johnson (1984), and Searle (1971), but these methods are not discussed in this section. A variety of statistical computer software packages are available that include algorithms for performing an analysis of variance with either balanced data or unbalanced and/or incomplete data. The use of such computerized algorithms is highly recommended.

Regardless of whether the α_i and β_j in Equation 5.186 are assumed to be fixed or random effects, the two-way crossed classification analysis of variance requires computation of the following quantities.

$$y_{ij} = \sum_{k=1}^n y_{ijk} \quad (5.187)$$

$$y_{i..} = \sum_{j=1}^b y_{ij} \quad (5.188)$$

$$y_{.j.} = \sum_{i=1}^a y_{ij} \quad (5.189)$$

$$y_{...} = \sum_{i=1}^a \sum_{j=1}^b y_{ij} \quad (5.190)$$

$$SS_{\text{Total}} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn} \quad (5.191)$$

$$SS_{\alpha} = \sum_{i=1}^a \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn} \quad (5.192)$$

$$SS_{\beta} = \sum_{j=1}^b \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn} \quad (5.193)$$

$$SS_{\text{Subclass}} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij}^2}{n} - \frac{y_{\dots}^2}{abn} \quad (5.194)$$

$$SS_{\alpha\beta} = SS_{\text{Subclass}} - SS_{\alpha} - SS_{\beta} \quad (5.195)$$

$$SS_{\epsilon} = SS_{\text{Total}} - SS_{\text{Subclass}} \quad (5.196)$$

$$df_{\text{Subclass}} = ab - 1 \quad (5.197)$$

$$df_{\alpha} = a - 1 \quad (5.198)$$

$$df_{\beta} = b - 1 \quad (5.199)$$

$$df_{\alpha\beta} = (a-1)(b-1) \quad (5.200)$$

$$df_{\epsilon} = ab(n-1) \quad (5.201)$$

$$MS_{\text{Subclass}} = SS_{\text{Subclass}}/df_{\text{Subclass}} \quad (5.202)$$

$$MS_{\alpha} = SS_{\alpha}/df_{\alpha} \quad (5.203)$$

$$MS_{\beta} = SS_{\beta}/df_{\beta} \quad (5.204)$$

$$MS_{\alpha\beta} = SS_{\alpha\beta}/df_{\alpha\beta} \quad (5.205)$$

$$MS_{\epsilon} = SS_{\epsilon}/df_{\epsilon} \quad (5.206)$$

When $n = 1$, the quantities SS_{ϵ} and df_{ϵ} are zero, and MS_{ϵ} is undefined. Thus, in the following sections, variance estimates and hypothesis tests that require these quantities are not possible unless $n > 1$.

5.4.2 The Fixed Effects Model

Assume that the levels of factors A and B are specifically chosen by the experimenter, and that all inferences and conclusions resulting from an analysis of the experimental data will be confined to these particular factor levels. In this case, the α_i , β_j , and $\alpha\beta_{ij}$ in Equation 5.186 are assumed to be fixed effects. The linear model given by Equation 5.186 can be rewritten as

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} \quad (5.207)$$

and it is assumed that y_{ijk} is a random observation from a population with mean μ_{ij} and with variance σ^2 for each combination of i and j . That is, changing the levels of factors A and B may cause the population mean to change, but it has no effect on the variance of the measured responses. For the purpose of testing hypotheses and constructing confidence intervals, it is assumed that $y_{ijk} \sim N(\mu_{ij}, \sigma^2)$.

The overall hypothesis that α_i , β_j , and $\alpha\beta_{ij}$ have no effect on the population mean is equivalent to the hypothesis that the subclass means μ_{ij} in Equation 5.207 are equal. This can be expressed formally as the hypotheses

$$H_0: \mu_{ij} = \mu_{i'j'} \text{ for all } i' \neq i \text{ and/or } j' \neq j \quad (5.208)$$

$$H_A: \mu_{ij} \neq \mu_{i'j'} \text{ for some } i' \neq i \text{ and/or } j' \neq j$$

Some authors, such as Bancroft (1968), advocate testing H_0 at the 0.25 level of significance. This is accomplished by computing the statistic

$$F = \frac{MS_{\text{Subclass}}}{MS_\epsilon} \quad (5.209)$$

where MS_{Subclass} and MS_ϵ are given by Equations 5.202 and 5.206, respectively. H_0 is rejected at the 0.25 level of significance if

$$F \geq F_{0.75}(df_{\text{Subclass}}, df_\epsilon) \quad (5.210)$$

where $F_{0.75}(df_{\text{Subclass}}, df_\epsilon)$ is obtained from Table A6. Note that this test is analogous to the one-way fixed effects analysis of variance test for equal class means presented in Section 5.1.1, where in this case, a class is a particular combination of the levels of factors A and B.

If H_0 is rejected, then the conclusion is that the μ_{ij} are probably not equal, which implies that some of the effects α_i , β_j , and $\alpha\beta_{ij}$ are not negligible, and thus, further testing is justified. The philosophy advocated by Bancroft (1968), however, is that if H_0 is not rejected at the 0.25 level of significance, then there is insufficient justification for further testing. It should be noted that testing at the 0.25 level of significance will result in detecting smaller differences among the subclass means than testing at, say, the 0.10 or 0.05 level. In fact, whether or not the μ_{ij} are unequal, testing at the 0.25 level of significance makes it more likely that H_0 will be rejected than if a smaller level of significance is used. Note also that it is quite possible in practice to reject the hypothesis of equal subclass means at the 0.25 level of significance, and then proceed to test the effects α_i , β_j , and $\alpha\beta_{ij}$ only to find that none of them is significant at a level smaller than 0.25.

Other authors, such as Scheffé (1959), Searle (1971), Hicks (1982), Kempthorne (1973), and Graybill (1976) consider directly testing the individual subhypotheses:

$$\begin{aligned} 1. H_0: \alpha_i &= 0, \text{ for all } i \\ H_A: \alpha_i &\neq 0, \text{ for some } i \end{aligned} \quad (5.211)$$

$$\begin{aligned} 2. H_0: \beta_j &= 0, \text{ for all } j \\ H_A: \beta_j &\neq 0, \text{ for some } j \end{aligned} \quad (5.212)$$

$$\begin{aligned} 3. H_0: \alpha\beta_{ij} &= 0, \text{ for all } i \text{ and } j \\ H_A: \alpha\beta_{ij} &\neq 0, \text{ for some } i \text{ and } j \end{aligned} \quad (5.213)$$

A useful step in testing these subhypotheses is the construction of an analysis of variance table, as illustrated in Table 5.10. From the E[MS] column of Table 5.10, note that if all $\alpha_i = 0$, then $E[MS_\alpha]$ reduces to σ^2 ; if all $\beta_j = 0$, then $E[MS_\beta]$ reduces to σ^2 ; and if all $\alpha\beta_{ij} = 0$, then $E[MS_{\alpha\beta}]$ reduces to σ^2 . If any of the effects α_i , β_j , or $\alpha\beta_{ij}$ are not zero, the corresponding mean squares will tend to overestimate σ^2 by the quantity shown in the E[MS] column. Thus, a large value of an F statistic in the F column of Table 5.10 indicates that the corresponding null hypothesis may be false.

Specifically, H_0 in Equation 5.211 is rejected at the α level of significance if

$$F = \frac{MS_\alpha}{MS_\epsilon} \geq F_{1-\alpha}(df_\alpha, df_\epsilon) \quad (5.214)$$

H_0 in Equation 5.212 is rejected at the α level of significance if

$$F = \frac{MS_\beta}{MS_\epsilon} \geq F_{1-\alpha}(df_\beta, df_\epsilon) \quad (5.215)$$

H_0 in Equation 5.213 is rejected at the α level of significance if

$$F = \frac{MS_{\alpha\beta}}{MS_\epsilon} \geq F_{1-\alpha}(df_{\alpha\beta}, df_\epsilon) \quad (5.216)$$

In these tests, the quantities $F_{1-\alpha}(df_\alpha, df_\epsilon)$, $F_{1-\alpha}(df_\beta, df_\epsilon)$, and $F_{1-\alpha}(df_{\alpha\beta}, df_\epsilon)$ are obtained from Table A.6. If any of the above

TABLE 5.10
Analysis of Variance Table for a Two-Way
Crossed Classification with a Fixed Effects Model

Source	df	SS	MS	E[MS]	F
Between levels of factor A	a-1	SS_{α}	MS_{α}	$\sigma^2 + \frac{bn}{a-1} \sum_{i=1}^a \alpha_i^2$	$MS_{\alpha}/MS_{\epsilon}$
Between levels of factor B	b-1	SS_{β}	MS_{β}	$\sigma^2 + \frac{an}{b-1} \sum_{j=1}^b \beta_j^2$	MS_{β}/MS_{ϵ}
Interaction of factors A and B	$(a-1)(b-1)$	$SS_{\alpha\beta}$	$MS_{\alpha\beta}$	$\sigma^2 + \frac{n}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta_{ij})^2$	$MS_{\alpha\beta}/MS_{\epsilon}$
Experimental error	$ab(n-1)$	SS_{ϵ}	MS_{ϵ}	σ^2	
Total	$abn-1$	SS_{Total}			

hypotheses are rejected, the next phase of the analysis involves estimation and comparison of means.

If the hypothesis $H_0: \alpha_i = 0$ is rejected, the implication is that the value of the population mean probably depends on the level of factor A. The nature of this dependence can be explored further by estimating the means

$$\mu_i = \mu + \alpha_i \quad (5.217)$$

The estimator of μ_i is the "row average" (referring to Table 5.9) given by

$$\bar{y}_{i.} = y_{i.}/bn \quad (5.218)$$

The variance of $\bar{y}_{i.}$ is

$$\text{Var}(\bar{y}_{i.}) = \sigma^2/bn \quad (5.219)$$

which is estimated by

$$\hat{\text{Vâr}}(\bar{y}_{i.}) = MS_\epsilon/bn \quad (5.220)$$

Note that $\hat{\text{Vâr}}(\bar{y}_{i.})$ is constant for all levels of factor A. Individual $100(1-\alpha)\%$ confidence intervals for the means μ_i are given by

$$\bar{y}_{i.} \pm \sqrt{MS_\epsilon/bn} \, t_{1-\alpha/2}(df_\epsilon) \quad (5.221)$$

where $t_{1-\alpha/2}(df_\epsilon)$ is obtained from Table A5. Simultaneous confidence intervals for the a means μ_i with an overall confidence level of $(1-\alpha)$ are constructed by applying Scheffé's method and are given by

$$\bar{y}_{i.} \pm \sqrt{(MS_\epsilon/bn)(a-1)F_{1-\alpha}(a-1, df_\epsilon)} \quad (5.222)$$

where $F_{1-\alpha}(b-1, df_\epsilon)$ is obtained from Table A6. Comparisons among the means are made by applying the procedures presented in Sections 5.1.4.2, 5.1.4.3, and 5.1.4.4. Note, however, that in all equations given in those sections, $\bar{y}_{i.}$ replaces \bar{y}_i , MS_ϵ and df_ϵ replace MS_{within} and $N - a$, and bn replaces n_i or n .

If the hypothesis $H_0: \beta_j = 0$ is rejected, the implication is that the value of the population mean depends on the level of factor B. This can be explored further by estimating the means

$$\mu_j = \mu + \beta_j \quad (5.223)$$

The estimator of μ_j is the "column average" (referring to Table 5.9) given by

$$\bar{y}_j = y_{.j}/an \quad (5.224)$$

The variance of \bar{y}_j is

$$\text{Var}(\bar{y}_j) = \sigma^2/an \quad (5.225)$$

which is estimated by

$$\text{V}\hat{\text{ar}}(\bar{y}_j) = \text{MS}_\epsilon/an \quad (5.226)$$

Note that $\text{V}\hat{\text{ar}}(\bar{y}_j)$ is constant over all levels of factor B. Individual $100(1-\alpha)\%$ confidence intervals for the means μ_j are given by

$$\bar{y}_j \pm \sqrt{\text{MS}_\epsilon/an} t_{1-\alpha/2}(df_\epsilon) \quad (5.227)$$

where $t_{1-\alpha/2}(df_\epsilon)$ is obtained from Table A5. Simultaneous confidence intervals for the b means μ_j with an overall confidence level of $(1-\alpha)$ are constructed by applying Scheffé's method and are given by

$$\bar{y}_j \pm \sqrt{(\text{MS}_\epsilon/an)(b-1)F_{1-\alpha}(b-1,df_\epsilon)} \quad (5.228)$$

where $F_{1-\alpha}(b-1,df_\epsilon)$ is obtained from Table A6. Comparisons among the means are made by applying the procedures presented in Sections 5.1.4.2, 5.1.4.3, and 5.1.4.4. Note that in the equations given in those sections, \bar{y}_j replaces \bar{y}_i , $b-1$ replaces $a-1$, MS_ϵ and df_ϵ replace $\text{MS}_{\text{within}}$ and $N-a$, and an replaces n_i or n .

When the hypothesis of no interaction is rejected, that is, when $H_0: \alpha\beta_{ij} = 0$ is rejected, interpretation of the effects of factors A and B may not be straightforward. This is because when factors A and B interact, the effects corresponding to the levels of factor A are different

for each level of factor B, and vice versa. Thus, comparisons among the row averages $\bar{y}_{i.}$ or the column averages $\bar{y}_{.j}$ may be meaningless or misleading. In such cases, the cell averages or subclass averages (referring to Table 5.9) must be examined.

Recall from Equation 5.207 that each observation y_{ijk} is from a population with mean μ_{ij} and variance σ^2 . The means μ_{ij} are estimated by the subclass averages

$$\bar{y}_{ij.} = y_{ij.}/n \quad (5.229)$$

The variance of $\bar{y}_{ij.}$ is

$$\text{Var}(\bar{y}_{ij.}) = \sigma^2/n \quad (5.230)$$

which is estimated by

$$\hat{\text{Var}}(\bar{y}_{ij.}) = MS_{\epsilon}/n \quad (5.231)$$

Individual $100(1-\alpha)\%$ confidence intervals for the means μ_{ij} are given by

$$\bar{y}_{ij.} \pm \sqrt{MS_{\epsilon}/n} t_{1-\alpha/2}(df_{\epsilon}) \quad (5.232)$$

where $t_{1-\alpha/2}(df_{\epsilon})$ is obtained from Table A5. Simultaneous confidence intervals for the ab means μ_{ij} , with an overall confidence level of $(1-\alpha)$, are constructed by using Scheffé's method and are given by

$$\bar{y}_{ij.} \pm \sqrt{(MS_{\epsilon}/n)(ab-1) F_{1-\alpha}(ab-1, df_{\epsilon})} \quad (5.233)$$

where $F_{1-\alpha}(ab-1, df_{\epsilon})$ is obtained from Table A6.

Comparisons can be made among the ab subclass averages using methods analogous to those presented in Sections 5.1.4.2, 5.1.4.3, and 5.1.4.4. One of the simplest and most effective methods of comparing the subclass means and interpreting interactions and factor effects, however, is to plot the subclass averages as discussed in Section 7.3.1. When constructing such plots, the scales of each axis should be carefully chosen. This is because lines and curves can be made to look nearly parallel or quite diverse, depending upon the scales of the axes. Also, for this same reason, such plots should not be used as the sole basis for determining whether or not interaction is present. The analysis of variance test given by

Equation 5.216 should be performed first. If the hypothesis $H_0: \alpha\beta_{ij} = 0$ is rejected, then plots of the subclass averages provide useful insight into the nature of the interaction. Then it may be desirable to apply Scheffé's method to test hypotheses about any notable patterns in the subclass averages.

5.4.3 The Random Effects Model

Consider the case where the levels of factors A and B are assumed to be randomly selected from larger populations of factor levels, and any conclusions from an analysis of the experimental data are to be extended to these populations. In this case, the observed response values in Table 5.9 are represented by the model in Equation 5.186. That is,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

$$i = 1, \dots, a$$

$$j = 1, \dots, b$$

$$k = 1, \dots, n$$

where α_i , β_j , and $\alpha\beta_{ij}$ are random effects. The terms in the random effects model are defined as follows:

- μ = the overall mean
- α_i = a random observation from a population with zero mean and with variance σ_α^2
- β_j = a random observation from a population with zero mean and with variance σ_β^2
- $\alpha\beta_{ij}$ = a random observation from a population with zero mean and with variance $\sigma_{\alpha\beta}^2$
- ϵ_{ijk} = a random observation from a population with zero mean and with variance σ^2 .

It is also assumed that the α_i , β_j , $\alpha\beta_{ij}$, and ϵ_{ijk} are mutually uncorrelated. Thus, each observation y_{ijk} is from a population with mean μ and with variance

$$\text{Var}(y_{ijk}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2 \quad (5.234)$$

For the purpose of hypothesis testing, it is further assumed that α_i , β_j , $\alpha\beta_{ij}$, and ϵ_{ijk} have normal distributions; thus, y_{ijk} is assumed to have a $N(\mu, \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2)$ distribution. This assumption, however, is not required for estimation of the variances or the overall mean μ .

The objectives of an analysis of variance with the random effects model are:

1. To test the hypotheses that $\sigma_\alpha^2 = 0$, $\sigma_\beta^2 = 0$, and $\sigma_{\alpha\beta}^2 = 0$.
2. To estimate the variance components σ_α^2 , σ_β^2 , and $\sigma_{\alpha\beta}^2$.
3. To estimate and perhaps test a hypothesis about the population mean μ .

After computing the quantities specified by Equations 5.187 through 5.206, the results are summarized in an analysis of variance table, such as Table 5.11. The hypotheses to be tested are

$$1. H_0 : \sigma_\alpha^2 = 0 \quad (5.235)$$

$$H_A : \sigma_\alpha^2 > 0$$

$$2. H_0 : \sigma_\beta^2 = 0 \quad (5.236)$$

$$H_A : \sigma_\beta^2 > 0$$

$$3. H_0 : \sigma_{\alpha\beta}^2 = 0 \quad (5.237)$$

$$H_A : \sigma_{\alpha\beta}^2 > 0$$

TABLE 5.11

**Analysis of Variance Table for a Two-Way
Crossed Classification with a Random Effects Model**

Source	df	SS	MS	E[MS]	F
Factor A	a-1	SS_α	MS_α	$\sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_\alpha^2$	$MS_\alpha/MS_{\alpha\beta}$
Factor B	b-1	SS_β	MS_β	$\sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2$	$MS_\beta/MS_{\alpha\beta}$
Interaction of factors A and B	(a-1)(b-1)	$SS_{\alpha\beta}$	$MS_{\alpha\beta}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$MS_{\alpha\beta}/MS_\epsilon$
Experimental error	ab(n-1)	SS_ϵ	MS_ϵ	σ^2	
Total	abn-1	SS_{Total}			

The expected mean square column of Table 5.11 provides guidance in determining which mean square should appear in the denominator of each F statistic. Note that if $\sigma_\alpha^2 = 0$, $E[MS_\alpha]$ is identical to $E[MS_{\alpha\beta}]$. That is, when $\sigma_\alpha^2 = 0$, MS_α and $MS_{\alpha\beta}$ are independent estimators for the same quantity. If $\sigma_\alpha^2 > 0$, however, then MS_α estimates the quantity given by $E[MS_\alpha]$ which is greater than $E[MS_{\alpha\beta}]$. Thus, the hypothesis H_0 in Equation 5.235 is rejected at the α level of significance if

$$F = \frac{MS_{\alpha}}{MS_{\alpha\beta}} \geq F_{1-\alpha}[a-1, (a-1)(b-1)] \quad (5.238)$$

Similarly, the hypothesis H_0 in Equation 5.236 is rejected at the α level of significance if

$$F = \frac{MS_{\beta}}{MS_{\alpha\beta}} \geq F_{1-\alpha}[b-1, (a-1)(b-1)] \quad (5.239)$$

Next, note that if $\sigma_{\alpha\beta}^2 = 0$, then $E[MS_{\alpha\beta}] = E[MS_{\epsilon}]$. If $\sigma_{\alpha\beta}^2 > 0$, however, then $E[MS_{\alpha\beta}] > E[MS_{\epsilon}]$. Thus, the hypothesis H_0 in Equation 5.237 is rejected at the α level of significance if

$$F = \frac{MS_{\alpha\beta}}{MS_{\epsilon}} \geq F_{1-\alpha}[(a-1)(b-1), ab(n-1)] \quad (5.240)$$

In these tests, the quantities $F_{1-\alpha}[\text{df}_{\text{numerator}}, \text{df}_{\text{denominator}}]$ are obtained from Table A6.

Estimates of the variance components σ_{α}^2 , σ_{β}^2 , $\sigma_{\alpha\beta}^2$, and σ^2 are computed by equating the MS quantities with the corresponding $E[MS]$ expressions and then solving the resulting four equations. This yields the estimators

$$\hat{\sigma}^2 = MS_{\epsilon} \quad (5.241)$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS_{\alpha\beta} - MS_{\epsilon}}{n} \quad (5.242)$$

$$\hat{\sigma}_{\beta}^2 = \frac{MS_{\beta} - MS_{\alpha\beta}}{an} \quad (5.243)$$

$$\hat{\sigma}_{\alpha}^2 = \frac{MS_{\alpha} - MS_{\alpha\beta}}{bn} \quad (5.244)$$

$\text{Var}(y_{ijk})$, given by Equation 5.234, is estimated by the quantity

$$\hat{\text{Vâr}}(y_{ijk}) = \hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\beta}^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2 \quad (5.245)$$

which has approximate degrees of freedom given by

$$\hat{df} = \frac{[\hat{\text{Var}}(y_{ijk})]^2}{\frac{\left[\left(\frac{n-1}{n}\right)MS_e\right]^2}{ab(n-1)} + \frac{\left[\left(\frac{ab-a-b}{abn}\right)MS_{\alpha\beta}\right]^2}{(a-1)(b-1)} + \frac{\left(\frac{MS_\alpha}{bn}\right)^2}{a-1} + \frac{\left(\frac{MS_\beta}{an}\right)^2}{b-1}} \quad (5.246)$$

where \hat{df} is rounded to the nearest integer.

If applying Equations 5.242 through 5.244 results in a negative estimate for any of the variance components, then the estimate is set to zero. This is logical, because a negative estimate could only result if the corresponding F statistic is less than one, in which case the corresponding null hypothesis is not rejected, and there is insufficient evidence to conclude that the particular variance component is larger than zero.

It is important to note that if $n = 1$, σ^2 and $\sigma_{\alpha\beta}^2$ cannot be separately estimated, and the null hypothesis that $\sigma_{\alpha\beta}^2 = 0$ cannot be tested, unless a prior estimate of σ^2 is available. With $n = 1$, however, it is still possible to estimate σ_α^2 and σ_β^2 and to test the hypotheses that they are zero.

An estimator for the overall mean μ is given by

$$\bar{y}_{...} = y_{...}/abn \quad (5.247)$$

which is the overall average of the observations. The variance of $\bar{y}_{...}$ is given by

$$\begin{aligned} \text{Var}(\bar{y}_{...}) &= \frac{bn\sigma_\alpha^2 + an\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 + \sigma^2}{abn} \\ &= \frac{E[MS_\alpha] + E[MS_\beta] - E[MS_{\alpha\beta}]}{abn} \end{aligned} \quad (5.248)$$

and is estimated by

$$\hat{\text{Var}}(\bar{y}_{...}) = \frac{MS_\alpha + MS_\beta - MS_{\alpha\beta}}{abn} \quad (5.249)$$

which has approximate degrees of freedom

$$\hat{df}^* = \frac{[abn\hat{\text{Var}}(\bar{y}_{...})]^2}{\frac{MS_\alpha^2}{a-1} + \frac{MS_\beta^2}{b-1} + \frac{MS_{\alpha\beta}^2}{(a-1)(b-1)}} \quad (5.250)$$

where \hat{df}^* is rounded to the nearest integer.

If it is assumed that $\bar{y}_{...}$ is from a normal distribution, then an approximate $100(1-\alpha)\%$ confidence interval for the overall mean μ is given by

$$\bar{y}_{...} \pm \sqrt{V\hat{a}r(\bar{y}_{...})} t_{1-\alpha/2}(df^*) \quad (5.251)$$

where $t_{1-\alpha/2}(df^*)$ is obtained from Table A5. Also, hypotheses about μ are tested by applying methods analogous to those presented in Section 4.2.2. The test statistic is given by

$$t = \frac{\bar{y}_{...} - \mu_0}{\sqrt{V\hat{a}r(\bar{y}_{...})}} \quad (5.252)$$

where μ_0 is a hypothesized value of μ , and t is compared with a value from Table A5 with degrees of freedom df^* .

5.4.4 The Mixed Effects Model

Consider now the situation where factor A has fixed effects and factor B has random effects. In this case, the model given by Equation 5.186 characterizes the observed response values, that is,

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \\ i &= 1, \dots, a \\ j &= 1, \dots, b \\ k &= 1, \dots, n \end{aligned}$$

where α_i is a fixed effect, but β_j and $\alpha\beta_{ij}$ are assumed to be random effects.

In presenting the mixed effects model, there are at least two different assumptions that can be made about the interaction effects $\alpha\beta_{ij}$. Both Searle (1971) and Hocking (1985) provide comparative discussions of two particular assumptions that result in different expected mean squares, variance component estimates, and denominators for the F statistic for testing the hypothesis that $\sigma_{\beta}^2 = 0$. As both authors point out, however, both assumptions or slight modifications of them appear frequently in statistical textbooks and literature. Typically, an author will adopt one assumption and will not mention any others. The approach selected for presentation in this chapter is consistent with a number of statistics texts, such as Ostle and Mensing (1979), Hicks (1982), and Snedecor and Cochran (1980), and should provide useful results that are reasonably sound for most applications. The choice of this particular approach, however, is not meant to imply that it is preferred over all others.

The terms in the mixed effects model are defined as follows:

- μ = the overall mean
- μ_i = $\mu + \alpha_i$ is the population mean for the i^{th} level of factor A; that is, α_i is a fixed effect
- β_j = a random observation from a population with zero mean and with variance σ_β^2
- $\alpha\beta_{ij}$ = a random observation from a population with zero mean and with variance $\sigma_{\alpha\beta}^2$
- ϵ_{ijk} = a random observation from a population with zero mean and with variance σ^2 .

Thus, each observation is assumed to be from a population with mean μ_i and with variance

$$\text{Var}(y_{ijk}) = \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2 \quad (5.253)$$

The objectives of the mixed effects analysis of variance are

1. To estimate and compare the means μ_i .
2. To estimate σ_β^2 , $\sigma_{\alpha\beta}^2$, and σ^2 .
3. To test the hypotheses that σ_β^2 and $\sigma_{\alpha\beta}^2$ are zero.

After applying Equations 5.187 through 5.206, the resulting quantities are used to construct an analysis of variance table, such as Table 5.12. The hypotheses to be tested are

$$1. H_0: \alpha_i = 0, \text{ for all } i \quad (5.254)$$

$$H_A: \alpha_i \neq 0, \text{ for some } i$$

$$2. H_0: \sigma_\beta^2 = 0 \quad (5.255)$$

$$H_A: \sigma_\beta^2 > 0$$

$$3. H_0: \sigma_{\alpha\beta}^2 = 0 \quad (5.256)$$

$$H_A: \sigma_{\alpha\beta}^2 > 0$$

The E[MS] column of Table 5.12 provides guidance in determining which mean square should appear in the denominator for each F statistic. Note that if $\alpha_i = 0$, for all i , then $E[MS_\alpha] = E[MS_{\alpha\beta}]$. That is, when all $\alpha_i = 0$, MS_α and $MS_{\alpha\beta}$ are independent estimators for the same quantity. If any $\alpha_i \neq 0$, however, then $E[MS_\alpha] > E[MS_{\alpha\beta}]$. Thus, $MS_{\alpha\beta}$ is the denominator of the F statistic, and the hypothesis H_0 in Equation 5.254 is rejected at the α level of significance if

$$F = \frac{MS_{\alpha}}{MS_{\alpha\beta}} \geq F_{1-\alpha}[a-1, (a-1)(b-1)] \quad (5.257)$$

Note that if $\sigma_{\beta}^2 = 0$, then $E[MS_{\beta}] = E[MS_{\epsilon}]$. Thus, MS_{ϵ} is the denominator of the F statistic, and the hypothesis H_0 in Equation 5.255 is rejected at the α level of significance if

$$F = \frac{MS_{\beta}}{MS_{\epsilon}} \geq F_{1-\alpha}[b-1, ab(n-1)] \quad (5.258)$$

TABLE 5.12

**Analysis of Variance Table for a Two-Way
Crossed Classification with a Mixed Effects
Model (α_i Fixed and β_j Random)**

Source	df	SS	MS	E[MS]	F
Factor A	$a-1$	SS_{α}	MS_{α}	$\sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{bn}{a-1} \sum_{i=1}^a \alpha_i^2$	$MS_{\alpha}/MS_{\alpha\beta}$
Factor B	$b-1$	SS_{β}	MS_{β}	$\sigma^2 + an\sigma_{\beta}^2$	MS_{β}/MS_{ϵ}
Interaction of factors A and B	$(a-1)(b-1)$	$SS_{\alpha\beta}$	$MS_{\alpha\beta}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$MS_{\alpha\beta}/MS_{\epsilon}$
Experimental error	$ab(n-1)$	SS_{ϵ}	MS_{ϵ}	σ^2	
Total	$abn-1$	SS_{Total}			

Similarly, the hypothesis H_0 in Equation 5.256 is rejected at the α level of significance if

$$F = \frac{MS_{\alpha\beta}}{MS_{\epsilon}} \geq F_{1-\alpha}[(a-1)(b-1), ab(n-1)] \quad (5.259)$$

Estimates of the variance components are obtained by equating MS_{β} , $MS_{\alpha\beta}$, and MS_{ϵ} with the corresponding $E[MS]$ expressions and then solving for σ^2 , $\sigma_{\alpha\beta}^2$, and σ_{β}^2 . This yields the estimators

$$\hat{\sigma}^2 = MS_{\epsilon} \quad (5.260)$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS_{\alpha\beta} - MS_{\epsilon}}{n} \quad (5.261)$$

$$\hat{\sigma}_\beta^2 = \frac{MS_\beta - MS_\epsilon}{an} \quad (5.262)$$

If application of Equation 5.260, 5.261, or 5.262 results in a negative estimate, the estimate is set to zero. The estimated variance of an individual observed value y_{ijk} is given by

$$\text{Vâr}(y_{ijk}) = \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}^2 \quad (5.263)$$

with approximate degrees of freedom given by

$$\hat{df} = \frac{[\text{anVâr}(y_{ijk})]^2}{\frac{MS_\beta^2}{b-1} + \frac{(aMS_{\alpha\beta})^2}{(a-1)(b-1)} + \frac{[(an-a-1)MS_\epsilon]^2}{ab(n-1)}} \quad (5.264)$$

If the hypothesis $H_0: \alpha_i = 0$ is rejected, it is of interest to estimate and compare the means μ_i corresponding to the levels of factor A. The estimator of μ_i is the row average (referring to Table 5.9) given by

$$\bar{y}_{i.} = y_{i.}/bn \quad (5.265)$$

The variance of $\bar{y}_{i.}$ is given by

$$\text{Var}(\bar{y}_{i.}) = \frac{n\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 + \sigma^2}{bn} \quad (5.266)$$

and is estimated by

$$\text{Vâr}(\bar{y}_{i.}) = \frac{MS_\beta - MS_\epsilon + aMS_{\alpha\beta}}{abn} \quad (5.267)$$

which has approximate degrees of freedom given by

$$\hat{df}^* = \frac{[abn \text{Vâr}(\bar{y}_{i.})]^2}{\frac{MS_\beta^2}{b-1} + \frac{(aMS_{\alpha\beta})^2}{(a-1)(b-1)} + \frac{MS_\epsilon^2}{ab(n-1)}} \quad (5.268)$$

Note that $\text{Vâr}(\bar{y}_{i.})$ and \hat{df}^* are constant for all levels of factor A.

Individual approximate $100(1 - \alpha)\%$ confidence intervals for the means μ_i are given by

$$\bar{y}_{i..} \pm \sqrt{\text{Vâr}(\bar{y}_{i..})} t_{1-\alpha/2}(\text{df}^*) \quad (5.269)$$

where $t_{1-\alpha/2}(\text{df}^*)$ is obtained from Table A5. Simultaneous confidence intervals for the a means μ_i with an overall $(1 - \alpha)$ confidence level are obtained by applying Scheffé's method, which gives

$$\bar{y}_{i..} \pm \sqrt{\text{Vâr}(\bar{y}_{i..}) (a-1) F_{1-\alpha}(a-1, \text{df}^*)} \quad (5.270)$$

where $F_{1-\alpha}(a-1, \text{df}^*)$ is obtained from Table A6.

Comparisons among the means are made by applying the procedures presented in Sections 5.1.4.2, 5.1.4.3, and 5.1.4.4. Note that in the equations given in those sections, $\bar{y}_{i..}$ replaces $\bar{y}_{i.}$, $MS_{\alpha\beta}$ and $(a - 1)(b - 1)$ replace MS_{Within} and $N - a$, and bn replaces n_i or n . This approach to comparing the means is only valid when the comparisons are contrasts of the means, where a contrast is defined by Equations 5.46 and 5.47.

REFERENCES

- Bancroft, T. A., 1968, *Topics in Intermediate Statistical Methods*, Iowa State University Press, Ames, Iowa.
- Bennett, C. A., and N. L. Franklin, 1954, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons, Inc., New York.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter, 1978, *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.
- Duncan, D. B., 1955, Multiple Range and Multiple F Tests, *Biometrics*, 11: 1-42.
- Graybill, F. A., 1976, *Theory and Application of the Linear Model*, Duxbury Press, Boston.
- Hicks, C. R., 1982, *Fundamental Concepts in the Design of Experiments*, 3rd Ed., Holt, Rinehart and Winston, Inc., New York.
- Hocking, R. R., 1985, *The Analysis of Linear Models*, Brooks/Cole Publishing Co., Monterey, California.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Kemphorne, O., 1973, *The Design and Analysis of Experiments*, Robert E. Krieger Publishing Co., Huntington, New York.
- Keuls, M., 1952, The Use of the Studentized Range in Connection with an Analysis of Variance, *Euphytica*, 1: 112-122.
- Miller, R., 1966, *Simultaneous Statistical Inference*, McGraw-Hill Publishing Co., New York.
- Milliken, G. A., and D. E. Johnson, 1984, *Analysis of Messy Data, Volume 1: Designed Experiments*, Lifetime Learning Publications, Belmont, California.
- Newman, D., 1939, The Distribution of the Range in Samples from a Normal Population Expressed in Terms of an Independent Estimate of Standard Deviation, *Biometrika*, 31: 20-30.

- Ostle, B., and R. W. Mensing, 1979, *Statistics in Research*, 3rd Ed., 2nd Printing, Iowa State University Press, Ames, Iowa.
- Pearson, E. S., and H. O. Hartley, 1951, Charts of the Power Function of the Analysis of Variance Tests, Derived from the Noncentral F-Distribution, *Biometrika*, **38**: 112-130.
- Petit, G. S., and C. A. Kienberger, 1961, Preparation of Stoichiometric U_3O_8 , *Anal. Chim. Acta*, **25**: 579-586.
- Satterthwaite, F. E., 1946, An Approximate Distribution of Estimates of Variance Components, *Biom. Bull.*, **2**: 110-114.
- Scheffé, H., 1959, *The Analysis of Variance*, John Wiley & Sons, Inc., New York.
- Searle, S. R., 1971, *Linear Models*, John Wiley & Sons, Inc., New York.
- , and R. F. Fawcett, 1970, Expected Mean Squares in Variance Components Models Having Finite Populations, *Biometrics*, **26**: 243-254.
- , and F. Pukelsheim, 1986, Effects of Intraclass Correlation on Weighted Averages, *The American Statistician*, **40**: 103-105.
- Snedecor, G. W., and W. G. Cochran, 1980, *Statistical Methods*, 7th Ed., Iowa State University Press, Ames, Iowa.
- Tukey, J. W., 1949, Comparing Individual Means in the Analysis of Variance, *Biometrics*, **5**: 99-114.

CHAPTER 6

Regression Analysis

6.0 INTRODUCTION

In nuclear material accounting, situations often arise where a variable of interest cannot readily be measured directly, but an estimate of its value can be inferred through relationships with one or more other variables that can be easily measured. For example, suppose that it is necessary to determine the volume of plutonium nitrate solution in a storage tank. Measuring this volume directly is difficult; however, the height of solution in the tank is more easily measured. Then, given the relationship between height and volume, the volume is obtained indirectly by measuring the height of solution in the tank. The development of a quantitative description of tank volume as a function of height is an example of a calibration problem.

The exact functional relationship between two or more variables is usually unknown. However, an approximating equation can often be developed by collecting data on the relevant variables across the range of interest and then fitting a function to the data. The process of postulating an appropriate functional form, estimating the parameters in this equation from the collected data, and evaluating the adequacy of the fitted equation is known as **regression analysis**.

As noted above, one or more variables related to the variable of interest can often be used to develop a regression equation. If the development is successful, these **independent variables** can then be used to estimate or predict a value for the variable of interest (**dependent variable**). Many nuclear material accounting applications of regression analysis use only one independent variable and one dependent variable, as in the above plutonium nitrate tank calibration problem. The primary applications of regression analysis in nuclear materials accounting pertain to calibration. Chapter 15 includes detailed presentations of several calibration problems.

6.1 AN OVERVIEW OF REGRESSION AND CORRELATION

In this chapter, **correlation** refers to the strength of a linear relationship between two variables.† **Regression** refers to the process of identifying

†The concept of correlation with more than two variables (multiple and partial correlation) does occur, but it is not discussed here. The discussion here is restricted to the correlation between two variables.

an approximating equation relating a dependent variable to one or more independent variables, estimating the unknown coefficients in the equation, and evaluating the adequacy of the resulting equation for predicting values of the dependent variable.

In this chapter, to be consistent with other texts on linear models and regression analysis, upper and lower case x and y are not used to distinguish between a random variable and an observed value. Instead, lower case x and y are used for both; whereas an upper case X with an underline (i.e., \underline{X}) is used to denote a matrix.

6.1.1 Correlation

The correlation (strength of a linear relationship) between two variables x and y is measured by the population correlation coefficient (Section 2.6), which is given by

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (6.1)$$

where σ_{xy} is the population covariance of x and y , and σ_x and σ_y are the population standard deviations of x and y . It can be shown that $-1 \leq \rho_{xy} \leq 1$,

where

$\rho_{xy} = 0$ indicates that x and y are uncorrelated,† i.e., there is no linear relationship between them

$\rho_{xy} = 1$ indicates that x and y have perfect positive correlation, i.e., all possible (x,y) pairs lie on a straight line with positive slope

$\rho_{xy} = -1$ indicates that x and y have perfect negative correlation, i.e., all possible (x,y) pairs lie on a straight line with negative slope

An estimator of ρ_{xy} is given by

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}}$$

†Note that two variables being uncorrelated does not imply that the variables are statistically independent. It means only that there is no linear relationship; a relationship of some other type may exist. However, if two variables are statistically independent, they are also uncorrelated.

or equivalently,

$$\hat{\rho}_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right]^{1/2} \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]^{1/2}} \quad (6.2)$$

where (x_i, y_i) is the i^{th} (\bar{x}, y) pair, and n is the number of (x, y) pairs in the sample. The estimator $\hat{\rho}_{xy}$ is called the sample correlation coefficient. However, in practice, the terms “correlation” and “correlation coefficient” are often used when referring to the sample correlation coefficient. That convention is followed in this book when no confusion results from doing so. Some example scatterplots of bivariate data [observed (x, y) pairs] and the corresponding sample correlation coefficients are displayed in Figure 6.1.

6.1.2 The Regression Process

The major steps in the regression process are outlined below. For simplicity, they are presented for the case where there is only one independent

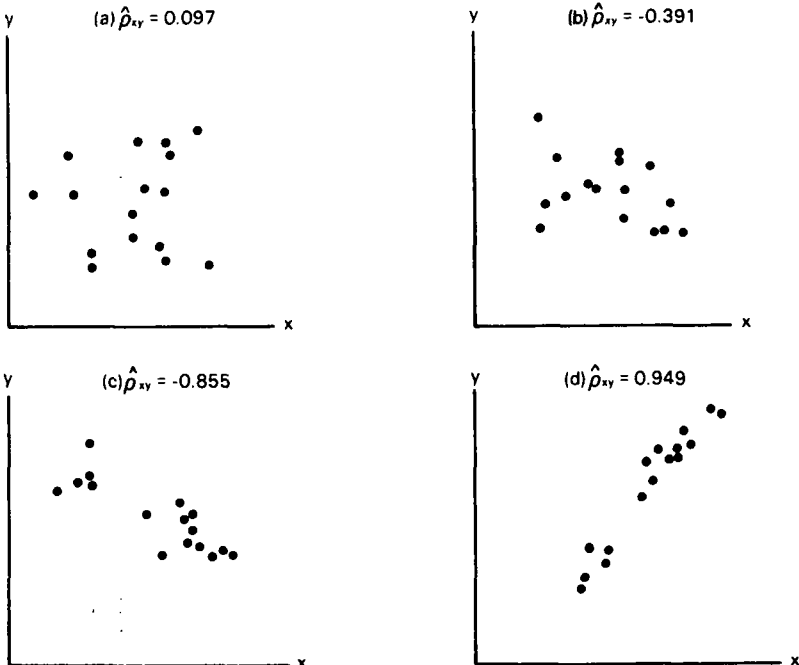


Figure 6.1 Example scatterplots with sample correlation coefficients.

variable. However, they readily extend to situations with more than one independent variable.

1. Formulate an experimental design for collecting the data to be used in developing the regression equation.† In designing such an experiment, a set of values of the independent variable must be chosen, and one or more values of the dependent variable are to be observed or measured at each value of the independent variable. The values of the independent variable should cover the range of interest and reflect the purpose of the analysis. The experiment must include at least as many distinct values of the independent variable as there are unknown parameters in the regression equation (see Steps 4 and 5). The experiment should also include the collection of more than one dependent variable observation at some of the independent variable values to provide an estimate of the experimental error variance. These replicate observations are of great value in assessing whether the fitted regression equation adequately represents the experimental data. Additional discussion of experimental design is given in Chapter 7.

2. Investigate the relationship between the dependent and independent variables with a scatterplot of the data. This could indicate, for example, whether a linear or curvilinear equation should be chosen to represent the unknown function.

3. Specify the measurement error structure inherent in the dependent and independent variables. In classical regression methods it is assumed that values of the independent variable are known without error and only the dependent variable is subject to error. While this assumption is seldom strictly valid, it is often reasonable because in many practical situations independent variable errors are much smaller than dependent variable errors. Regression techniques do exist for the case where there are significant errors in both the dependent and independent variables; however, they are somewhat more complicated (see Section 6.2.4.3). Other error structure questions of concern include the following:

a. Are the errors in the dependent variable statistically dependent; i.e., will a given observation tend to have a large (or small) error if the previous observation had a large (or small) error, or will successive errors tend to have the same sign?

b. Do the dependent variable errors have a constant variance over the independent variable range of interest, or does the error variance change with the value of the observation?

Specifying the error structure is very important in determining the correct model fitting (parameter estimation) technique in Step 5.

†A designed experiment is not a necessity for a regression analysis; often in practice, only data from undesigned experiments are available. However, without a well-designed set of experimental data, the results of a regression analysis may be inconclusive and/or inadequate.

4. Write down the mathematical model that represents the form of the regression equation indicated in Step 2. This model should include the error structure specified in Step 3 and the parameters to be estimated in Step 5. For example, suppose that variables x and y are approximately linearly related, values of x are assumed to be known without error, and values of y are subject to errors that are independent and identically normally distributed. Then an appropriate mathematical model is given by

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where

- y_i = an observed value of y at $x = x_i$
- β_0, β_1 = unknown regression parameters (coefficients) to be estimated from the (x, y) data
- ϵ_i = error associated with y_i ; the $\epsilon_i, i = 1, 2, \dots, n$, are assumed to be independent and identically normally distributed with mean zero and variance σ^2

5. Choose the appropriate model fitting technique (ordinary, weighted, or nonlinear least squares) and estimate the parameters (coefficients) in the regression equation. For the linear example above, the estimated parameters are $\hat{\beta}_0$ and $\hat{\beta}_1$, where the hat (^) indicates an estimate of the unknown true parameter. The least squares estimation techniques are discussed in detail later in this chapter.

6. Check the adequacy of the fitted regression equation obtained in Step 5 and the underlying assumptions. This should include an analysis of variance, a test for lack of fit, and an analysis of residuals (a residual is the difference between an observed value of the dependent variable and the corresponding value predicted from the fitted model). These analyses aid in the detection of outlying or influential data points and can indicate whether a modification of the form of the regression equation is required.

7. If necessary, modify the form of the regression equation or perform a transformation of the dependent variable and go to Step 4. Note that if a dependent variable transformation is performed, the error structure will likely change.

8. When the final regression equation is obtained, quantify the error introduced by using the regression equation to predict (indirectly measure) the dependent variable based on knowledge of the independent variable.†

The above steps are presented as a brief overview of the regression process and are discussed in more detail in subsequent sections of this chapter.

†In calibration problems, the dependent variable in the regression equation usually becomes the independent variable in application (see Chapter 15).

This overview is intended to provide an appreciation of the many necessary steps in a thorough regression analysis. In practice, however, many of these steps are often incorrectly overlooked.

In performing a regression analysis, remember that the hypothesized regression model is often just a model that is tentatively adopted because the true equation is unknown. In fitting the model to the collected data, however, it is implicitly assumed that the regression model is correct; thus, this assumption must be checked. This check may result in either adopting the tentative model as adequate or rejecting it (with more work required).

6.2 SIMPLE LINEAR REGRESSION BASICS

Simple linear regression refers to a regression between one dependent and one independent variable, where the relationship between the two variables is a straight line. This section uses an example with a simple error structure to introduce and illustrate some of the concepts presented in the regression steps of Section 6.1.2. Detailed discussions of these steps for other error structures are presented in subsequent sections.

6.2.1 An Example

The following example is adopted from Jaech (1973). The data are from an undesigned experiment, and there are no replicate observations as suggested in Step 1 of the previous section. The data in this example are used throughout this chapter to illustrate the various estimation and hypothesis testing methods.

Example 6.1 The following data consist of measurements for total ^{235}U made by an NDA instrument on 1-gal containers known to contain specified amounts of ^{235}U .

$x = \text{g of } ^{235}\text{U}$	$y = \text{net counts}$
10	890
15	1234
20	1491
25	1815
40	2896
50	3718

The dependent variable (y) and independent variable (x) are defined as:

y = net counts produced by the NDA instrument
 x = known amount of ^{235}U in a given container

It is assumed that the random error variance associated with the dependent variable y is constant over the range of the data. Although this assumption may not be realistic for all situations that involve counting data (i.e., counting data are often from a Poisson distribution where the variance is equal to the mean), it is made in this example for simplicity of illustration. A regression equation with y as a function of x is to be developed.

Example 6.1 is actually a special type of regression problem known as a calibration problem. A relationship between amount of ^{235}U and counts on the NDA instrument is required so that the amount of ^{235}U may be measured indirectly with the NDA instrument. The regression equation is obtained by fixing values of x and then observing y , but it is actually implemented by observing y and predicting the x value that corresponds to it. This section presents methods for developing a regression equation for y as a function of x . The calibration applications of regression are discussed in detail in Chapter 15.

A scatterplot of the data from Example 6.1 is displayed in Figure 6.2. Although there is an indication of slight curvature, it appears that a

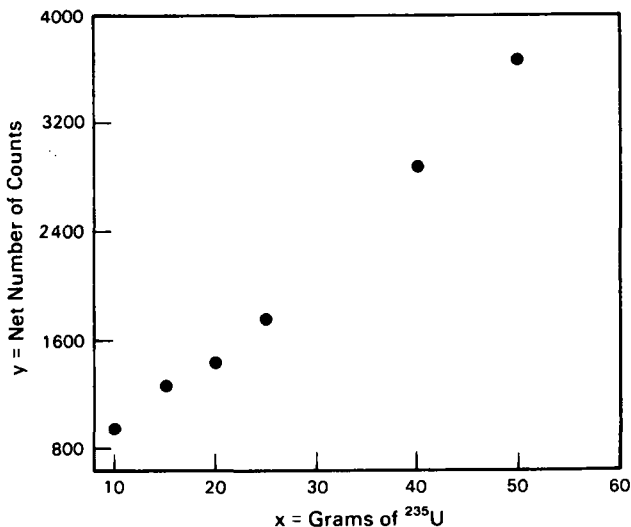


Figure 6.2 Scatterplots of data from Example 6.1.

straight line might adequately explain the relationship between grams of ^{235}U and net counts on the NDA instrument.

6.2.2 Model and Error Structure

The error structure for Example 6.1 is specified in the problem description. The amount of ^{235}U in each container (x) is assumed to be known without error (or the error is relatively small compared to the error in the number of counts). The number of counts (y) is subject to error; the errors are independent and have a constant variance over the range of interest.

The mathematical model, representing a hypothesized linear relationship and a specified error structure, is given by

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (6.3)$$

where

- y_i = observed net counts for container i
- β_0 = true unknown intercept of the line
- β_1 = true unknown slope of the line
- x_i = ^{235}U content of container i in g
- ϵ_i = random error in y_i

For subsequent discussions, the following assumptions are made:

1. ϵ_i is a normally distributed random variable with mean zero and variance σ^2 .
2. ϵ_i and ϵ_j , $i \neq j$, are independent (and hence uncorrelated). Thus, $E[y_i] = \beta_0 + \beta_1 x_i$ and $\text{Var}[y_i] = \sigma^2$.

While these assumptions are often reasonable, they should not be arbitrarily applied. Statistical procedures for examining these assumptions are considered in Section 6.4. However, nonstatistical consideration of the error structure inherent in the system plays a significant role in formulating correct assumptions.

So far, Steps 1 through 4 of the regression process outlined in Section 6.1 have been considered. The next step is the selection of a method for estimating the parameters β_0 and β_1 .

6.2.3 Ordinary Least Squares

It would be most unusual if all the data points (x_i, y_i) in a scatterplot, such as Figure 6.2, fell exactly on a straight line. Generally, if a straight line is drawn through the points in a scatterplot† the points will vary about

†This means that a straight line is drawn such that the data points are not either all above or all below the line. It is possible that none of the data points fall exactly on the line. In such a case, at least one data point must be above and at least one data point must be below the line.

the line, but most of them will not fall on the line. It is desirable to choose a line through the data points that minimizes the variability of the points about the line. The **least squares** criterion provides a technique for doing so.

Figure 6.3 is an enlarged portion of Figure 6.2 with a line drawn through the data points (x_i, y_i) . Each value of x (x_i) has a corresponding

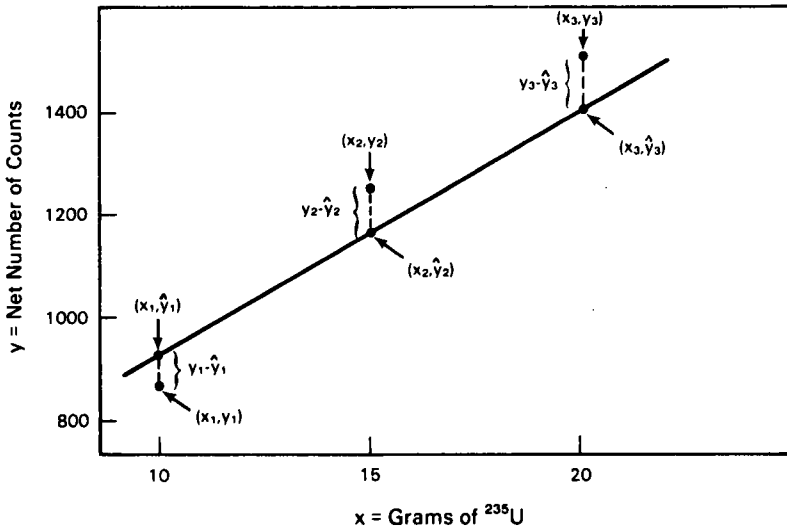


Figure 6.3 An enlarged portion of Figure 6.1 showing residuals.

predicted y value (\hat{y}_i) lying on the line and an observed y value (y_i). The quantity $y_i - \hat{y}_i$ is referred to as the i^{th} **residual**; hence,

$$SS_{\text{Residual}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

is called the **residual sum of squares**, where n is the number of data points (x_i, y_i) . The least squares criterion defines the line that results in the smallest value of SS_{Residual} . The line associated with the smallest residual sum of squares is called the **least squares line**. Note that the least squares criterion is valid for fitting any equation where y is expressed as a function of one or more independent variables; it is not restricted to straight lines.

Parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained from the least squares criterion by minimizing SS_{Residual} , where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ for a linear equation; that is,

$$SS_{\text{Residual}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

minimized with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ by calculus techniques. This is discussed in greater detail in most texts on linear models or regression analysis, such as Graybill (1976) or Draper and Smith (1981). The resulting estimators are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (6.4)$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (6.5)$$

where

$$\bar{x} = \sum_{i=1}^n x_i / n \quad \text{and} \quad \bar{y} = \sum_{i=1}^n y_i / n$$

Note that when the errors ϵ_i are normally distributed, the least squares estimators are also maximum likelihood estimators.

Example 6.2 Using the data from Example 6.1, fit a simple linear regression equation relating y (net number of counts) to x (g of ^{235}U).

The following intermediate results are obtained from the data given in Example 6.1:

$$\begin{array}{ll} \sum_{i=1}^n x_i = 160 & \sum_{i=1}^n y_i = 12,044 \\ \sum_{i=1}^n x_i^2 = 5450 & \sum_{i=1}^n y_i^2 = 30,042,502 \\ \sum_{i=1}^n x_i y_i = 404,345 & n = 6 \end{array}$$

Then the coefficient estimates are computed from Equations 6.4 and 6.5, which give

$$\hat{\beta}_1 = \frac{6(404,345) - (160)(12,044)}{6(5450) - (160)^2} = 70.286$$

and

$$\hat{\beta}_0 = \frac{12,044}{6} - 70.286 \left(\frac{160}{6} \right) = 133.040$$

and the fitted least squares regression line is

$$\hat{y} = 133.040 + 70.286 x$$

Before a fitted regression equation is used to make predictions for unobserved data points, it should be evaluated to determine how well it fits the observed data. Regression is an iterative procedure that should continue until any model inadequacies or data anomalies are resolved. Evaluation techniques are discussed in Section 6.4. First, however, other regression models are introduced and discussed, with a general development in Section 6.3.

6.2.4 Other Regression Models

Often in regression situations with one independent variable, a straight line will not adequately represent the unknown true relationship. Likewise, when there is more than one independent variable, an equation linear in the independent variables may not adequately represent the unknown true relationship. Sometimes curvilinear models are needed that incorporate algebraic functions of the independent variables. The independent variables, and any new variables generated by functions of them, are called **predictor variables**. A regression problem with more than one predictor variable is called a **multiple regression** problem.

The other aspect of a regression model (besides the functional form) is the assumed error structure. Generally, the independent variables are assumed to be known without error. This class of regression problems (including various assumptions about the error structure of the dependent variable) is considered in detail in Section 6.3. However, there are situations where both independent and dependent variables are subject to considerable error. This situation is discussed briefly in Section 6.2.4.3.

6.2.4.1 Simple and Multiple Regression Equations

A wide variety of functional forms is used to formulate regression models. For illustrative purposes, a list of several common equations is given in Table 6.1.

TABLE 6.1
Several Common Regression Equations

Name	Equation
1. Hyperbola	$y = \frac{1}{\beta_0 + \beta_1 x}$ or $\frac{1}{y} = \beta_0 + \beta_1 x$
2. Exponential	$y = \beta_0 \beta_1^x$ or $\log y = \log \beta_0 + (\log \beta_1)x$
3. Geometric	$y = \beta_0 x^{\beta_1}$ or $\log y = \log \beta_0 + \beta_1 \log x$
4. Quadratic	$y = \beta_0 + \beta_1 x + \beta_2 x^2$
5. Cubic	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$
6. Linear—2 independent variables	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
7. Quadratic—2 independent variables	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$

Some nonlinear regression equations can be fitted using simple linear regression techniques by employing algebraic transformations, as in the first three equations in Table 6.1. Equations of this type are said to be intrinsically linear. The fourth and fifth equations in Table 6.1 represent members of the class of polynomial equations with one independent variable. The last two equations illustrate polynomial equations with more than one independent variable. While the equations in Table 6.1 by no means include all useful forms, they do illustrate some commonly used forms of simple and multiple regression equations.

6.2.4.2 Transformations of the Dependent Variable

In regression analysis, where ordinary least squares is used to estimate the regression equation parameters, the usual error assumptions are that the errors are independently and identically normally distributed with mean zero and a common variance σ^2 . The assumption of normally distributed errors is not necessary for least squares estimation of the parameters. However, when the errors in the dependent variable are not normally distributed, the ability to make inferences from the regression analysis is restricted. When the errors do not have a common variance, ordinary least squares should not be used to estimate the parameters (one option is to

use weighted least squares; see Section 6.3.2). In both of these situations—nonnormally distributed errors or heterogeneous error variances—it is sometimes possible to obtain approximately normal errors or homogeneous error variances through transformations of the dependent variable.

Transformations to stabilize several types of heterogeneous variance situations are displayed below. If a transformation is appropriate, y' (rather than y) is used as the dependent variable.

Variance proportional to $E[y]$: $y' = \sqrt{y}$ or $y' = \sqrt{y + 1}$

Standard deviation proportional to $E[y]$: $y' = \log y$

Standard deviation proportional to $(E[y])^2$: $y' = \frac{1}{y}$

For a more detailed discussion of these and other transformations, see Draper and Smith (1981).

In practice, nonnormality and heterogeneous variances often occur simultaneously. Fortunately, the transformations that help stabilize variances are often useful in correcting for nonnormality. It is a wise policy, however, to check the residuals (discussed in Section 6.4.5) to make sure the transformation is effective in both stabilizing variances and providing an approximately normal distribution of the errors.

6.2.4.3 Errors in Both Variables

Consider a simple linear regression where both the dependent and independent variables are subject to error. As in Equation 6.3, the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (6.6)$$

is assumed, but now

$$x'_i = x_i + e_i \quad (6.7)$$

where the x'_i are fixed by the experimenter, x_i is the true value of x'_i , and e_i is the random error associated with x_i . By substituting Equation 6.7 into Equation 6.6, the model becomes

$$y_i = \beta_0 + \beta_1 x'_i + (\epsilon_i - \beta_1 e_i) \quad (6.8)$$

with error structure assumptions $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, $e_i \sim N(0, \sigma_e^2)$, and $\text{Cov}(\epsilon_i, e_i) = 0$. Then Equation 6.8 can be written as

$$y_i = \beta_0 + \beta_1 x_i' + \epsilon_i^* \quad (6.9)$$

where $\epsilon_i^* = \epsilon_i - \beta_1 \epsilon_i$ and $\epsilon_i^* \sim N(0, \sigma_\epsilon^2 + \beta_1^2 \sigma_\epsilon^2)$

This model is similar in both form and error structure to the model in Equation 6.3, and hence β_0 and β_1 are estimated by ordinary least squares as in Section 6.2.3. For more information on this model, see Berkson (1950) or Draper and Smith (1981). Applications of this model to calibration are considered in Chapter 15.

In the multiple regression situation where each of several independent variables is subject to error, the discussion of regression becomes more complicated (see Seber, 1977).

6.3 GENERAL MULTIPLE LINEAR REGRESSION

Simple linear regression is introduced in Section 6.2, where it is assumed that the errors are independent with a constant variance. The possibility of other error structure assumptions is mentioned, as are many other simple and multiple regression equations that might be used. This section presents a general framework for a large class of regression problems.

6.3.1 The General Linear Regression Model

Recall that a regression model is composed of an assumed functional form and an error structure. Consider the class of models in which the predictor variables x_1, x_2, \dots, x_{p-1} are known without error, and the functional form is linear in the regression parameters; that is,

$$y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij} + \epsilon_i \quad (6.10)$$

This class of models can be represented in a general form by using matrix notation.† This **general linear regression model** is given by

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon} \quad (6.11)$$

where

\underline{y} = an $n \times 1$ vector of observable random variables
 \underline{X} = an $n \times p$ matrix ($n > p$) of known fixed numbers
 $\underline{\beta}$ = a $p \times 1$ vector of unknown regression parameters

†Readers unfamiliar with matrix notation should refer to a text on matrix algebra, such as Graybill (1969), Searle (1966), or Searle and Hausman (1970).

$\underline{\epsilon}$ = an $n \times 1$ vector of unobservable random variables (representing errors in y)

$$E[\underline{\epsilon}] = \underline{0}$$

$$\text{Cov}[\underline{\epsilon}] = \underline{\Sigma} \text{ (the variance-covariance matrix)}$$

Note that Equation 6.11 is Equation 6.10 written in matrix notation for $i = 1, 2, \dots, n$; that is,

$$\begin{array}{c} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ n \times 1 \end{array} = \begin{array}{c} \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{bmatrix} \\ n \times p \end{array} \begin{array}{c} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \\ p \times 1 \end{array} + \begin{array}{c} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \\ n \times 1 \end{array}$$

The variance-covariance matrix $\underline{\Sigma}$ is

$$\text{Cov}(\underline{\epsilon}) = \underline{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

where σ_i^2 is the variance of ϵ_i , $i = 1, 2, \dots, n$, and σ_{ij} is the covariance of ϵ_i and ϵ_j (see Section 2.6.3).

The general linear model given by Equation 6.11 has many special cases, depending upon the error structure assumptions. The following two special cases are of particular interest in nuclear material control applications:

Case 1: $\underline{\epsilon} \sim N(0, \sigma^2 \underline{I})$

σ^2 unknown and ϵ_i , $i = 1, 2, \dots, n$, are independent (\underline{I} is the $n \times n$ identity matrix)

Case 2: $\underline{\epsilon} \sim N(0, \sigma^2 \underline{V})$

σ^2 unknown and \underline{V} is an $n \times n$ matrix of known constants (if the ϵ_i are independent, then \underline{V} is a diagonal matrix)

Estimation of the regression parameter vector β for each of these cases is considered in the following section.

6.3.2 Parameter Estimation

As discussed in Section 6.2.3, the least squares criterion requires minimization of the residual sum of squares, given by

$$SS_{\text{Residual}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

or in matrix notation

$$(\underline{y} - \hat{\underline{y}})'(\underline{y} - \hat{\underline{y}}) \quad (6.12)$$

where $\hat{\underline{y}} = \underline{X}\hat{\beta}$. The least squares estimators for β for Cases 1 and 2 are presented below.

Case 1: Ordinary Least Squares

$$\hat{\beta} = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y} \quad (6.13)$$

Case 2: Weighted Least Squares

Here, the ordinary least squares estimator given by Equation 6.13 does not apply directly. Instead, estimates are obtained by transforming the observations \underline{y} to new variables \underline{z} that meet the Case 1 assumptions; that is,

$$\underline{z} = \underline{P}^{-1}\underline{y} = \underline{P}^{-1}\underline{X}\beta + \underline{P}^{-1}\underline{\epsilon} \quad (6.14)$$

or

$$\underline{z} = \underline{Q}\beta + \underline{f} \quad (6.15)$$

where

$$\underline{P}^{-1} = \text{inverse of the symmetric matrix } \underline{P} \text{ where } \underline{P} \underline{P}' = \underline{V}$$

$$\underline{Q} = \underline{P}^{-1} \underline{X}$$

$$\underline{f} = \text{random vector with } E[\underline{f}] = 0 \text{ and } \text{Cov}[\underline{f}] = \sigma^2 \underline{I}$$

Then the estimator for β is given by

$$\hat{\beta} = (Q'Q)^{-1}Q'z = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y} \quad (6.16)$$

A common weighted least squares (Case 2) application occurs when the errors are assumed to be independent but they do not have a constant variance. In this situation, the matrix \underline{V} is diagonal, so that the variance-covariance matrix is

$$\text{Cov}[\underline{\epsilon}] = \sigma^2\underline{V} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \quad (6.17)$$

where $\text{Var}[\epsilon_i] = \sigma_i^2$. Because the σ_i^2 are seldom known, estimates are substituted in practical applications. Estimates may be obtained from prior information or from replicate observations. If estimates of the σ_i^2 cannot be obtained prior to the regression analyses, it may be necessary to make the erroneous assumption that $\underline{V} = \underline{I}$, proceed with the regression analysis, and then attempt to learn something about the form of \underline{V} by examining the residuals. Note that the application of ordinary least squares to a situation where weighted least squares should be used will yield unbiased coefficient estimates, but the estimates are not minimum variance. For a more detailed presentation of weighted least squares regression, see Draper and Smith (1981). Graybill (1976) provides considerable theoretical coverage of the general linear regression model.

Matrix notation has been adopted in the literature to simplify the general representation of regression models and the associated computations in a regression analysis. Note that a change in the number of predictor variables or observations does not change the matrix notation. However, when there are more than one or two predictor variables, the required computations become tedious and impractical to do by hand. Computer packages that perform these regression calculations are widely available, but care should be exercised in choosing a package to ensure that it is appropriate for the application at hand.

6.4 EVALUATING A FITTED REGRESSION EQUATION

This section introduces statistical techniques for evaluating how well a fitted regression equation represents a given set of data. The techniques

presented are for multiple regression in general but are illustrated by using the data from the simple linear regression problem presented in Example 6.1.

6.4.1 Regression Analysis of Variance

As discussed in Chapter 5, analysis of variance (ANOVA) methods are applied to study the sources of variability that affect a dependent variable in a set of data by partitioning the total sum of squares into components that correspond to the sources of variability. For regression, the partitioning can be written as

$$\left(\begin{array}{c} \text{Total sum} \\ \text{of squares} \end{array} \right) = \left(\begin{array}{c} \text{Sum of squares} \\ \text{explained by the regression} \end{array} \right) + \left(\begin{array}{c} \text{Residual sum} \\ \text{of squares} \end{array} \right)$$

or

$$SS_{\text{Total}} = SS_{\text{Reg}} + SS_{\text{Residual}}$$

For Case 1 (Section 6.3.1) this is given by

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6.18)$$

Computational formulas for the components in Equation 6.18 are given by

$$SS_{\text{Total}} = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n}$$

$$SS_{\text{Reg}} = \hat{\beta}' \underline{X}' \underline{y} - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n}$$

and

$$SS_{\text{Residual}} = SS_{\text{Total}} - SS_{\text{Reg}}$$

For the simple linear regression model, given by Equation 6.3, the formula for SS_{Reg} reduces to the form

$$\begin{aligned}
 SS_{\text{Reg}} &= \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
 &= \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \frac{\left[\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n} \right]^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n}}
 \end{aligned}$$

For Case 1, Equation 6.18 partitions the total variability in the observed y_i values into one component that is accounted for by the fitted regression model and one that the fitted model does not explain. From this partitioning, a statistic that quantifies the usefulness of the fitted regression equation is given by

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Total}}} \quad (6.19)$$

This statistic is referred to as the **coefficient of determination** and represents the fraction of the total variability in the observed y_i explained by the fitted regression model. The closer R^2 is to 1, the better the fitted regression model represents the observed data.

Any sum of squares (SS) has associated with it a number called its degrees of freedom (df). This number indicates how many independent pieces of information involving y_1, y_2, \dots, y_n are used to compute a sum of squares. From the components in Equation 6.18 an ANOVA table is constructed, such as the one displayed in Table 6.2 for a regression equation with p parameters. (For a simple linear regression with an intercept term, $p = 2$.) The mean squares (MS) are obtained by dividing each sum of squares by its corresponding degrees of freedom.

The residual mean square, denoted by s^2 , provides an estimate of the variance about the fitted regression model, based on $n - p$ degrees of freedom. This variance includes components due to both pure error (experimental error observed from replicates) and the lack of fit of the regression

TABLE 6.2

**Analysis of Variance (ANOVA) Table for a Regression
Equation with p Parameters**

Source	df	SS	MS
Regression	$p - 1$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{SS_{\text{reg}}}{p - 1}$
Residual	$n - p$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$s^2 = \frac{SS_{\text{Residual}}}{n - p}$
Total	$n - 1$	$\sum_{i=1}^n (y_i - \bar{y})^2$	

equation fitted to the data. Statistical techniques for testing lack of fit are considered in Section 6.4.4.

6.4.2 F-Test for Significance of Regression

While the R^2 statistic provides a quantitative measure of how much variability is explained by a fitted regression model, it does not indicate whether the amount explained is statistically significant. The question of statistical significance is answered by testing the hypothesis

$$H_0: \beta_i = 0, \text{ for all } i=1, \dots, p - 1$$

against the alternative

$$H_A: \text{at least one } \beta_i \neq 0, i=1, \dots, p - 1$$

When H_0 is true, the test statistic

$$F = \frac{MS_{\text{Reg}}}{s^2} \quad (6.20)$$

has an F-distribution with $(p - 1)$ and $(n - p)$ degrees of freedom (the F-distribution is introduced in Section 3.5.1). If $F \geq F_{1-\alpha}(p - 1, n - p)$, found in Table A6, the null hypothesis is rejected; that is, the fitted regression equation accounts for a statistically significant portion of the total variability. This implies that H_A is true; that is, at least one of the $\beta_i \neq 0$. Note that the test statistic in Equation 6.20 is only valid for testing the above hypothesis when there is no significant lack of fit (see Section 6.4.4).

Example 6.3 The regression analysis that began in Examples 6.1 and 6.2 is continued by calculating the quantities in the ANOVA table and the coefficient of determination R^2 . Determine whether the fitted regression line in Example 6.2 explains a significant portion of the variability. Use the $\alpha = 0.05$ level of significance.

By using the quantities computed in Example 6.2, the component sums of squares in Equation 6.18 are computed as

$$SS_{\text{Total}} = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = 30,042,502 - \frac{(12,044)^2}{6} = 5,866,179.33$$

$$SS_{\text{Reg}} = \frac{\left[\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}\right]^2}{\left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}\right]} = \frac{\left[404,345 - \frac{(160)(12,044)}{6}\right]^2}{\left[5,450 - \frac{(160)^2}{6}\right]} = 5,845,796.73$$

and

$$SS_{\text{Residual}} = SS_{\text{Total}} - SS_{\text{Reg}} = 20,382.60$$

The resulting ANOVA table is given below.

Source	df	SS	MS
Regression	1	5,845,796.73	5,845,796.73
Residual	4	20,382.60	$s^2 = 5,095.65$
Total	5	5,866,179.33	

The test statistic, computed from Equation 6.20, is

$$F = \frac{MS_{\text{Reg}}}{s^2} = \frac{5,845,796.73}{5,095.65} = 1,147.21$$

The value $F_{0.95}(1,4) = 7.71$ is obtained from Table A6. Because F exceeds 7.71, the null hypothesis $H_0: \beta_1 = 0$ is rejected in favor of the alternative $H_A: \beta_1 \neq 0$ (i.e., the fitted regression equation accounts for a statistically significant portion of the total variability).

By substituting quantities from the ANOVA table into Equation 6.19, the coefficient of determination is

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Total}}} = \frac{5,845,796.73}{5,866,179.33} = 0.997$$

Thus, the fitted linear regression equation explains 99.7% of the variability in the observed y values.

6.4.3 Sampling Distribution of $\hat{\beta}$

Because regression parameters are estimated from experimental or sampling data, the least squares estimator $\hat{\beta}$ for the unknown coefficient vector has a sampling distribution. For either Case 1 or Case 2, the elements of the vector $\hat{\beta}$ (the individual estimators $\hat{\beta}_i$) are jointly distributed (see Section 2.6.3) and have a multivariate normal distribution. It was noted in Section 6.3.2 that $\hat{\beta}$, whether defined by Equation 6.13 for Case 1 or by Equation 6.16 for Case 2, is an unbiased estimator for the unknown coefficient vector β , provided that the specified form of the regression model is correct. If so, the distribution of $\hat{\beta}$ has a mean vector $E[\hat{\beta}] = \beta$.

For Case 1, where $\text{Cov}[\epsilon] = \sigma^2 I$ (see Section 6.3.1), $\hat{\beta}$ has a p -variate normal distribution with mean vector β and $p \times p$ variance-covariance matrix given by

$$\text{Cov}[\hat{\beta}] = \sigma^2 (X'X)^{-1}$$

$$= \begin{bmatrix} \text{Var}[\hat{\beta}_0] & \text{Cov}[\hat{\beta}_0, \hat{\beta}_1] \cdots & \text{Cov}[\hat{\beta}_0, \hat{\beta}_{p-1}] \\ \text{Cov}[\hat{\beta}_0, \hat{\beta}_1] & \text{Var}[\hat{\beta}_1] & \vdots \\ \vdots & \ddots & \text{Cov}[\hat{\beta}_{p-2}, \hat{\beta}_{p-1}] \\ \text{Cov}[\hat{\beta}_0, \hat{\beta}_{p-1}] \cdots & \text{Cov}[\hat{\beta}_{p-2}, \hat{\beta}_{p-1}] & \text{Var}[\hat{\beta}_{p-1}] \end{bmatrix} \quad (6.21)$$

For the simple linear regression model with unknown coefficients β_0 and β_1 , Equation 6.21 has the form

$$\text{Var}[\hat{\beta}] = \sigma^2(\underline{X}'\underline{X})^{-1} = \begin{bmatrix} \text{Var}[\hat{\beta}_0] & \text{Cov}[\hat{\beta}_0, \hat{\beta}_1] \\ \text{Cov}[\hat{\beta}_0, \hat{\beta}_1] & \text{Var}[\hat{\beta}_1] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{-\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} \\ \frac{-\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix}$$

That is,

$$\text{Var}[\hat{\beta}_0] = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\text{Cov}[\hat{\beta}_0, \hat{\beta}_1] = \frac{-\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

When σ^2 is unknown, the variances and covariances are estimated by substituting s^2 for σ^2 in Equation 6.21, when it is appropriate to do so (s^2 is an unbiased estimator for σ^2 only when the fitted equation does not have a significant lack of fit; see Section 6.4.4). Thus, for the simple linear

regression model, estimators for the variances and covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\begin{aligned} \text{Cov}[\hat{\beta}] &= s^2(\underline{X}'\underline{X})^{-1} \\ &= \begin{bmatrix} \frac{s^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{-s^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} \\ \frac{-s^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} & \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix} \end{aligned} \quad (6.22)$$

That is,

$$\text{Var}[\hat{\beta}_0] = \frac{s^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\text{Var}[\hat{\beta}_1] = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\text{Cov}[\hat{\beta}_0, \hat{\beta}_1] = \frac{-s^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

For Case 2, where $\text{Cov}[\underline{\epsilon}] = \sigma^2 \underline{V}$ (see Section 6.3.1), $\hat{\beta}$ has a p -variate normal distribution with mean vector $\underline{\beta}$ and $p \times p$ variance-covariance matrix given by

$$\text{Cov}[\hat{\beta}] = \sigma^2(\underline{X}'\underline{V}^{-1}\underline{X})^{-1} \quad (6.23)$$

When σ^2 is unknown, the appropriate estimator for Case 2 is given by

$$\hat{\sigma}^2 = \frac{1}{n - p} \underline{y}' [\underline{Y}^{-1} - \underline{Y}^{-1} \underline{X} (\underline{X}' \underline{Y}^{-1} \underline{X})^{-1} \underline{X}' \underline{Y}^{-1}] \underline{y} \quad (6.24)$$

which is then substituted for σ^2 in Equation 6.23.

For either Case 1 or Case 2, $\hat{\beta}$ has a multivariate normal distribution and each element $\hat{\beta}_i$ has a marginal distribution that is normal with mean β_i and variance $\text{Var}[\hat{\beta}_i]$. Thus, the significance of the individual coefficient estimates $\hat{\beta}_i$ can be investigated by testing the hypothesis

$$H_0: \beta_i = 0$$

against the alternative

$$H_A: \beta_i \neq 0$$

When H_0 is true, the test statistic

$$t_i = \frac{\hat{\beta}_i}{\sqrt{\hat{\text{Var}}[\hat{\beta}_i]}} \quad (6.25)$$

has a Student's t -distribution with $(n - p)$ degrees of freedom. Thus, if $|t_i| \geq t_{1-\alpha/2}(n - p)$, found in Table A5, then H_0 is rejected and the parameter estimate is said to be significantly different from zero, which implies that the true value of β_i is probably different from zero.

Care must be taken in applying these tests to the individual coefficient estimates because of the correlations among the estimates. In practice (assuming that more than one coefficient has been estimated), only the coefficient estimate with the smallest t statistic is a candidate for deletion from the regression model. If this candidate coefficient is found to be non-significant, the model is refitted with only the remaining coefficients being estimated. At this point, the individual tests are applied again, and the least significant coefficient estimate is eliminated. This process continues until all remaining coefficient estimates (if any) are statistically significant. Because the estimates are usually correlated, however, removing one coefficient from the model can cause others that have already been removed to become significant. This would not be detected if the above procedure were implemented. Stepwise regression methods such as those discussed in Draper and Smith (1981) are often used to select an adequate submodel from a hypothesized full model form.

Methods are available for simultaneously testing the significance of several coefficient estimates. The most common method uses a full vs. reduced model concept [see Draper and Smith (1981) for details of this method].

Example 6.4 Continuing with the regression analysis started in Examples 6.1, 6.2, and 6.3, test whether $\hat{\beta}_0$ and $\hat{\beta}_1$ are different from zero at the $\alpha = 0.10$ level of significance.

By applying Equation 6.22 with $s^2 = MS_{\text{Residual}}$ from Example 6.3, the variance estimates are computed as

$$\widehat{\text{Var}}[\hat{\beta}_0] = \frac{5095.65 (5450)}{6(1183.33)} = 3911.4607$$

and

$$\widehat{\text{Var}}[\hat{\beta}_1] = \frac{5095.65}{1183.33} = 4.3062$$

For each coefficient estimate, a test statistic is computed from Equation 6.25, which gives

$$t_0 = \frac{\hat{\beta}_0}{\sqrt{\widehat{\text{Var}}[\hat{\beta}_0]}} = \frac{133.040}{\sqrt{3911.4607}} = 2.13$$

and

$$t_1 = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{Var}}[\hat{\beta}_1]}} = \frac{70.286}{\sqrt{4.3062}} = 33.87$$

With $\alpha = 0.10$, the value $t_{0.95}(4) = 2.13$ is obtained from Table A5. Because $t_0 = 2.13$ and $t_1 > 2.13$, $\hat{\beta}_0$ and $\hat{\beta}_1$ are significantly greater than zero. The significance of $\hat{\beta}_0$ is interesting. It is logical in this example to expect β_0 to be zero; that is, $x = 0$ g of ^{235}U should produce zero counts on the NDA instrument if the background adjustment is made correctly. Because $\hat{\beta}_0$ is significantly greater than zero, perhaps the NDA instrument was incorrectly adjusted for background. Another possible explanation is that the assumed model is incorrect and $\hat{\beta}_0$ is biased by the effect of one or more predictor variables not incorporated in the model.

6.4.4 Testing for Lack of Fit

During a regression analysis, a regression equation is fitted to the data, based on an assumed model form. This fitted equation should not be

blindly accepted as the correct model but only tentatively considered. If the true model is substantially different from the assumed model, then the fitted model should have a significant **lack of fit**. This lack of fit can be investigated by partitioning the residual sum of squares into its component parts; that is,

$$\left(\begin{array}{c} \text{Residual sum} \\ \text{of squares} \end{array} \right) = \left(\begin{array}{c} \text{Pure error} \\ \text{sum of squares} \end{array} \right) + \left(\begin{array}{c} \text{Lack of fit} \\ \text{sum of squares} \end{array} \right)$$

or

$$SS_{\text{Residual}} = SS_{\text{PE}} + SS_{\text{LOF}} \quad (6.26)$$

Note that a significant lack of fit will inflate the residual sum of squares and hence, the residual mean square. If a prior estimate † of σ^2 (the pure error variance) is available, it can be compared with the residual mean square by performing an F-test. If the residual mean square is significantly greater than the prior estimate of σ^2 , there is a significant lack of fit and the model is deemed inadequate.

If no prior estimate of σ^2 exists but replicate observations‡ have been taken on y at one or more values of x , these replicates can be used to compute an estimate of σ^2 . Because such an estimate is often more reliable for the current problem at hand than a prior estimate, provisions for replicate observations should be made when designing the experiment that produces the data. Experimental design is the topic of Chapter 7, and the reader is referred there for additional details.

Additional notation is required when replicate observations are present in the data. Consider a simple linear regression: suppose that there are m different values of x , and at the i^{th} of these values, x_i , there are $n_i > 1$ replicate observations. Then

$y_{11}, y_{12}, \dots, y_{1n_1}$ are n_1 replicate observations at x_1

$y_{21}, y_{22}, \dots, y_{2n_2}$ are n_2 replicate observations at x_2

\vdots \vdots \vdots

$y_{m1}, y_{m2}, \dots, y_{mn_m}$ are n_m replicate observations at x_m

†A prior estimate is one obtained from previous experience with the situation being studied.

‡Replicate observations must be genuine repeats and not just repetitions of a reading or measurement. When collecting the data, replicates should be randomly interspersed with and treated as any other observations (see Chapter 7 for details).

where

$$n = \sum_{i=1}^m n_i$$

is the total number of observations on y . The pure error sum of squares is obtained by pooling the corrected sum of squares of the y_{ij} at each x_i ; that is,

$$SS_{PE} = \sum_{i=1}^m \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2 \quad (6.27)$$

where

$$\bar{y}_i = \sum_{k=1}^{n_i} y_{ik}/n_i$$

This sum of squares has degrees of freedom given by

$$df_{PE} = \sum_{i=1}^m (n_i - 1) = n - m$$

Thus, the pure error mean square is

$$MS_{PE} = \frac{SS_{PE}}{df_{PE}} \quad (6.28)$$

and is an unbiased estimator for σ^2 (the common variance of the errors) irrespective of whether or not the assumed model form is correct.

Subtracting the pure error sum of squares from the residual sum of squares yields the lack of fit sum of squares; that is,

$$SS_{LOF} = SS_{Residual} - SS_{PE} \quad (6.29)$$

which has degrees of freedom

$$df_{LOF} = df_{Residual} - df_{PE} = m - p$$

When replicate observations are available, the information in Equations 6.26 through 6.29 can be used to create an expanded ANOVA table as shown in Table 6.3. The degrees of freedom for this table are based upon

TABLE 6.3
Expanded ANOVA Table

Source	df	SS	MS
Regression	$p - 1$	$\sum_{i=1}^m \sum_{k=1}^{n_i} (\hat{y}_{ik} - \bar{y})^2$	$\frac{SS_{\text{reg}}}{p - 1}$
Residual	$n - p$	$\sum_{i=1}^m \sum_{k=1}^{n_i} (y_{ik} - \hat{y}_{ik})^2$	$\frac{SS_{\text{Residual}}}{n - p}$
Lack of fit	$m - p$	$SS_{\text{Residual}} - SS_{\text{PE}}$	$\frac{SS_{\text{LOF}}}{m - p}$
Pure error	$n - m$	$\sum_{i=1}^m \sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2$	$\frac{SS_{\text{PE}}}{n - m}$
Total	$n - 1$	$\sum_{i=1}^m \sum_{k=1}^{n_i} (y_{ik} - \bar{y})^2$	

a model with p predictor variables (including the constant term), and n_i replicates at the i^{th} set of predictor values, for $i = 1, 2, \dots, m$.

An F-test for lack of fit is performed by computing the test statistic

$$F = \frac{MS_{\text{LOF}}}{MS_{\text{PE}}} \quad (6.30)$$

and comparing it with $F_{1-\alpha}(m - p, n - m)$ found in Table A6. If F is greater than or equal to the table value, it indicates that the assumed model form does not adequately represent the true model. Otherwise, the assumed model is accepted as adequate, and the residual mean square can be used as an estimate of σ^2 .

Note that the pure error sum of squares in Equation 6.27 is equivalent to the within class sum of squares given by Equations 5.16 and 5.20 for the one-way ANOVA. To apply the ANOVA methods from Chapter 5, each distinct set of values of the predictor variables is treated as a class. (In Example 6.5 below, each value of x is treated as a class, so there are $m = 6$ classes.) Then the between class sum of squares given by Equations 5.15 and 5.19 includes all variations in y that are accounted for by varying the levels of the predictor variables. This includes the regression sum of squares and the lack of fit sum of squares; that is,

$$SS_{\text{Between}} = SS_{\text{Reg}} + SS_{\text{LOF}}$$

Thus, the necessary components for a lack of fit test can be obtained from computer software packages by making two passes through the data. The first pass is a regression analysis which yields SS_{Reg} , SS_{Residual} , and their degrees of freedom. The second pass is a one-way ANOVA which yields SS_{Between} , SS_{PE} , and their degrees of freedom. These quantities are then used to fill in the ANOVA table as shown in Table 6.3.

Example 6.5 Consider the problem described in Example 6.1, but assume that two independent counts were taken at each x_i value as follows:

$x = \text{g of } ^{235}\text{U}$	$y = \text{net counts}$
10	890, 905
15	1234, 1216
20	1491, 1511
25	1815, 1798
40	2896, 2886
50	3718, 3740

Fit a simple linear regression model to the 12 data points and determine whether there is significant lack of fit at the $\alpha = 0.05$ level.

A scatterplot of the 12 data points is displayed in Figure 6.4. Although there is a hint of curvature, a simple linear equation of the form $y = \beta_0 + \beta_1 x$ is fitted to the data. By applying Equations 6.4, 6.5, 6.23, and 6.24, the least squares coefficient estimates and their estimated variances are computed as

$$\begin{aligned} \hat{\beta}_0 &= 131.39 & \hat{\beta}_1 &= 70.39 \\ \widehat{\text{Var}}[\hat{\beta}_0] &= 1812.2049 & \widehat{\text{Var}}[\hat{\beta}_1] &= 1.9881 \end{aligned}$$

By applying Equation 6.25, the test statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$t_0 = \frac{131.39}{42.57} = 3.09 \quad \text{and} \quad t_1 = \frac{70.39}{1.41} = 49.92$$

Both values exceed $t_{0.975}(10) = 2.23$; that is, both estimated parameters are significantly greater than zero at the $\alpha = 0.05$ level.

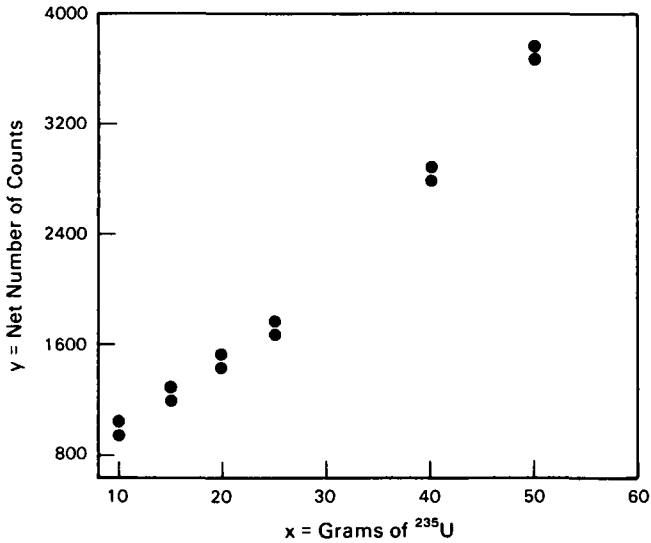


Figure 6.4 Scatterplot of replicated NDA instrument calibration data.

The expanded analysis of variance table, with quantities computed according to Table 6.3, is given by

Source	df	SS	MS
Regression	1	11,724,646	11,724,646
Residual	10	47,219	4,721.9
Lack of fit	4	46,308	11,577.0
Pure error	6	911	151.8
Total	11	11,771,865	

The test statistic for lack of fit is computed from Equation 6.27, which gives

$$F = \frac{MS_{\text{LOF}}}{MS_{\text{PE}}} = \frac{11,577.0}{151.8} = 76.3$$

Because F is larger than $F_{0.95}(4,6) = 4.53$, the fitted simple linear regression model exhibits a significant lack of fit. Note that even though the

simple linear regression model explains 99.6% of the variability in y (i.e., $R^2 = 0.996$), the unexplained variability cannot be attributed solely to experimental (counting) error. The conclusion is that a lack of fit exists and the equation $\hat{y} = 131.39 + 70.39x$ does not adequately fit the data. Thus, the assumed linear model given by Equation 6.3 may not be appropriate; perhaps some other model form should be considered, depending on the desired prediction accuracy.

A significant lack of fit implies that the assumed model form is inadequate, and a new model form must be specified. Residual analysis techniques discussed in the next section are helpful in choosing appropriate modifications of the assumed model. As with any hypothesis test, a nonsignificant lack of fit does not prove that the assumed model is adequate. The interpretation of a nonsignificant lack of fit depends upon the size of the experimental error variance σ^2 relative to the size of deviations from the true model form that would be of concern to the practitioner. The power of the test is also affected by the number of degrees of freedom for lack of fit and for pure error. However, with these considerations in mind, if a nonsignificant lack of fit is interpreted as implying that the hypothesized model form is adequate, then residual analysis is still appropriate to check the error structure assumptions of the model.

6.4.5 Examining Residuals

Residuals are defined as the differences

$$r_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n \quad (6.31)$$

where y_i is an observation and \hat{y}_i is the associated predicted value obtained from the fitted regression equation. Intuitively, the residuals are the parts of the observations that the fitted regression equation does not explain. Thus, for a fitted regression equation with a significant lack of fit, examination of the residuals may suggest appropriate modifications to the model form. For a regression equation with no significant lack of fit, each r_i may be thought of as an observed random error if the model form is correct. Because the form of the equation is assumed to be correct in this case, the residuals can be examined to see whether they support the assumed error structure of the model. For example, a commonly assumed error structure has independent errors with zero mean and a constant variance σ^2 , and the errors are assumed to have a normal distribution. The residuals should tend to confirm these assumptions, or at least should not contradict them.

6.4.5.1 Plotting Residuals

Plotting the residuals can be very helpful in checking the validity of the assumed model. Some useful ways of plotting the residuals r_i (or the standardized residuals discussed in Section 6.4.5.2) are:

1. In a histogram or probability plot.
2. Against the fitted values \hat{y}_i .
3. Against the independent variables x_{ij} .
4. In observation order, if known.

Histogram or probability plot. If an assumed model is correct, the histogram of residuals should resemble a normal distribution with mean zero.† An alternative is to form a normal probability plot of the residuals and then observe whether the points fall approximately on a straight line. The probability plot is preferred over the histogram when the number of data points is not large. Both the histogram and the probability plot require a subjective judgment; however, Filliben (1975) presents an objective test based upon the normal probability plot. Other tests for normality are discussed in Chapter 9.

Residuals vs. \hat{y}_i . Several possible patterns that might be observed in a scatterplot of the r_i 's vs. the corresponding \hat{y}_i 's are shown in Figure 6.5. These plots indicate, respectively, that the following conditions prevail:

1. Satisfactory behavior of the residuals (no anomalies).

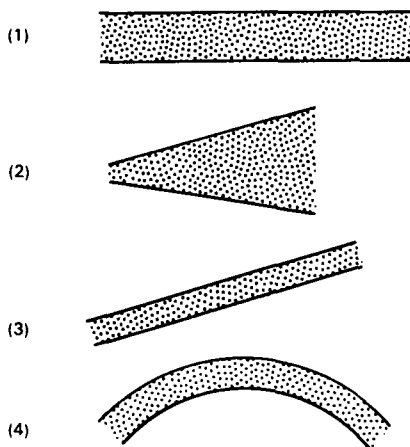


Figure 6.5 Example residual scatterplot patterns.

†Note that the average of the residuals will be zero if the regression model includes a constant term (β_0).

2. Variance of the errors is not constant. For this example, the variance increases as the value of the dependent variable increases. Either application of weighted least squares or a transformation of the dependent variable is needed (see Sections 6.3 and 6.2.4.2).

3. Error in analysis. This can be caused by incorrectly omitting from the model either a constant term or a linear term involving an independent variable.

4. The model form is inadequate. Curvature terms (squares and/or cross products of independent variables) should be added to the model. Alternatively, there may be a need for a transformation of the x_{ij} and/or y_i before performing the regression (see Sections 6.2.4.1 and 6.2.4.2).

Note that combinations and variations of the anomalies 2, 3, and 4 can occur in practice.

Residuals vs. independent variables. Here a scatterplot is formed for each independent variable x_j by plotting the r_i 's against the corresponding x_{ij} 's. The possible patterns illustrated in Figure 6.5 and discussed above still apply. These plots accommodate an examination of the model with respect to each independent variable and sometimes provide guidance in correcting deficiencies due to a particular x_j .

Residuals vs. observation order. Plotting the residuals r_i against their observation order is helpful in detecting unsuspected trends or variations in the data over time. Again, referring to Figure 6.5, the horizontal pattern 1 suggests no time effect, while patterns 2, 3, and 4 would indicate unaccounted for time effects in the error variance or the form of the model.

Other residual plots. If a regression analysis yields a fitted equation with a significant lack of fit and the above residual plots do not help resolve the problem, other residual plots may be helpful. It is often enlightening to plot the residuals against other variables that were not controlled during the experimentation. For example, observations may have been obtained with several supposedly similar measuring instruments, but there may be an instrument effect that could help explain some of the previously unexplained variability in y . In general, residuals should be plotted in any reasonable way that the experimenter or data analyst feels might provide some insight into the cause of nonrandom or unusual patterns in the residuals. For additional techniques and more detailed discussions, see Cook and Weisberg (1982) and Daniel and Wood (1980).

Summary. The subjective residual plotting procedures presented above are quite useful; in applications, their subjectiveness seldom hinders their utility. However, many objective statistical analysis techniques exist for the examination of residuals. Some of these are presented in Section 6.4.6. The interested reader may also want to refer to Anscombe (1961), Anscombe and Tukey (1963), or some of the other references given in Section 6.4.6.

Example 6.6 Continuing Example 6.5, analyze the residuals and modify the regression equation accordingly.

The residuals and other pertinent information are displayed in Table 6.4. Plots of the residuals (r_i) against the predicted values (\hat{y}_i) and against the independent variable (x_i) are displayed in Figures 6.6 and 6.7, respectively. The pattern of the residuals is very similar in the two plots, because there is only one independent variable in this example. Each scatterplot shows a definite curvature indicating that additional terms should be added to the model to account for this curvature; thus, an x_i^2 term is considered.

TABLE 6.4

Data with Predicted Values and Residuals from the Fitted Simple Linear Regression Model

i	x_i	y_i	\hat{y}_i	r_i
1	10	890	835.3	54.7
2	10	905	835.3	69.7
3	15	1234	1187.2	46.8
4	15	1216	1187.2	28.8
5	20	1491	1539.0	-48.0
6	20	1511	1539.0	-28.0
7	25	1815	1891.0	-76.0
8	25	1798	1891.0	-93.0
9	40	2896	2946.8	-50.8
10	40	2886	2946.8	-60.8
11	50	3718	3650.7	67.3
12	50	3740	3650.7	89.3

The expanded model form is given by

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_{ij}$$

for $i = 1, 2, \dots, 6$ and $j = 1, 2$. By using the least squares algorithm in a statistical computer software package, the coefficient estimates, their estimated standard deviations (i.e., the square roots of their estimated variances), and test statistics are as follows:

	$\hat{\beta}_i$	$\sqrt{\hat{\text{Var}}[\hat{\beta}_i]}$	$t_i = \hat{\beta}_i / \sqrt{\hat{\text{Var}}[\hat{\beta}_i]}$
β_0	419.83	30.95	13.56
β_1	45.665	2.46	18.56
β_2	0.408	0.04	10.20

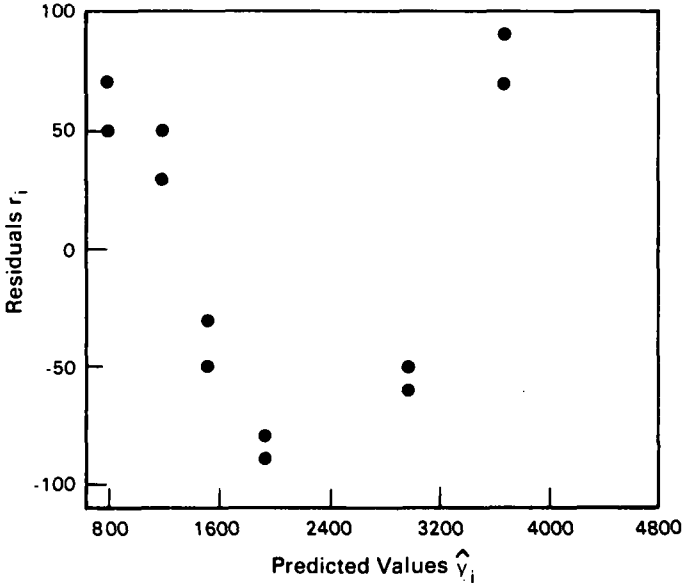


Figure 6.6 Scatterplot of residuals vs. predicted values for the simple linear regression model.

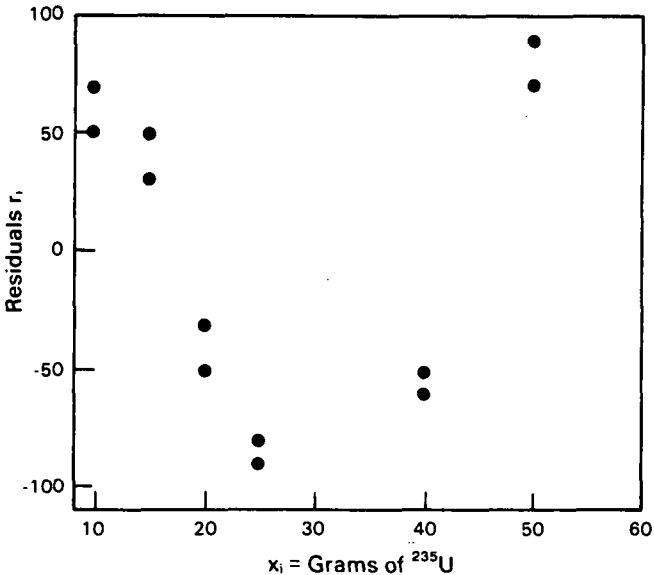


Figure 6.7 Scatterplot of residuals vs. values of the independent variable for the simple linear regression model.

The t_i statistics indicate that all coefficient estimates are significantly greater than zero [e.g., for $\alpha = 0.01$, $t_{0.995}(9) = 3.25$]. The expanded analysis of variance table, computed according to Table 6.3, is given below.

Source	df	SS	MS
Regression	2	11,768,113	5,884,056
Residual	9	3,752	416.9
Lack of fit	3	2,841	947.0
Pure error	6	911	151.8
Total	11	11,771,865	

The test statistic for lack of fit, computed from Equation 6.30, is

$$F = \frac{947.0}{151.8} = 6.24$$

If the $\alpha = 0.05$ level of significance is chosen, the value $F_{0.95}(3,6) = 4.76$ is obtained from Table A6. Because $6.24 > 4.76$, the lack of fit is significant at the $\alpha = 0.05$ level. Note, however, that the lack of fit is much less pronounced than in the simple linear model fitted in Example 6.5.

Although the fitted quadratic model has statistically significant lack of fit, the “practical significance” should be carefully considered before proceeding with additional modifications of the model form. With $R^2 = 0.9996$, a statistically significant lack of fit may be of no real practical importance when the requirements of the intended use of the model are considered. Note that the pure error mean square, MS_{PE} , indicates the limitation on the repeatability of the observed y_{ij} values at a given x_i value. For discussion purposes, note that

$$\frac{\sqrt{MS_{PE}}}{\bar{y}} = \frac{12.32}{2008.33} = 0.006$$

That is, the estimated pure error standard deviation is about 0.6% of the overall average of the observed y values. Note also that if the lack of fit is ignored and the residual mean square is used as an estimate of the experimental error variance, then the estimated relative standard deviation is

$$\frac{\sqrt{MS_{Residual}}}{\bar{y}} = \frac{20.42}{2008.33} = 0.010$$

or 1.0%. That is, including the lack of fit deviations increases the estimated relative standard deviation from 0.6% to 1.0%. Depending upon the intended use of the fitted model, this may be acceptable. However, omitting the x^2 term from the model inflates the estimated relative standard deviation to 3.4%, which is almost six times the estimated relative pure error standard deviation. Again, this may be acceptable for some applications; but for this example, we conclude that the linear model is inadequate and use the quadratic model.

The observed x and y values, the predicted values, and the residuals for the fitted quadratic model are displayed in Table 6.5. A plot of the residuals from the fitted quadratic equation against x is shown in Figure 6.8. (Because of its similarity to this plot, the plot of r_i against \hat{y}_i is not given here.)

TABLE 6.5
Data with Predicted Values and Residuals from
the Fitted Quadratic Regression Model

i	x_i	y_i	\hat{y}_i	r_i
1	10	890	917.3	-27.3
2	10	905	917.3	-12.3
3	15	1234	1196.7	37.3
4	15	1216	1196.7	19.3
5	20	1491	1496.4	-5.4
6	20	1511	1496.4	14.6
7	25	1815	1816.6	-1.6
8	25	1798	1816.6	-18.6
9	40	2896	2899.5	-3.5
10	40	2886	2899.5	-13.5
11	50	3718	3723.5	-5.5
12	50	3740	3723.5	16.5

There is an indication of an effect due to a higher power of x (as indicated by the rough curve drawn through the points in Figure 6.8). However, when a third-order polynomial is fitted to the data, addition of the cubic term does not significantly reduce the lack of fit. In fact, a fourth-order polynomial is required if the lack of fit must be reduced to a level that is not statistically significant. This, along with the above considerations, suggests that for practical purposes the fitted quadratic model adequately represents the unknown true model form.

The next step involves checking the error structure assumptions. The residuals plotted in Figure 6.8 do not appear to contradict the assumption of a constant variance; however, with only two observed y values at each

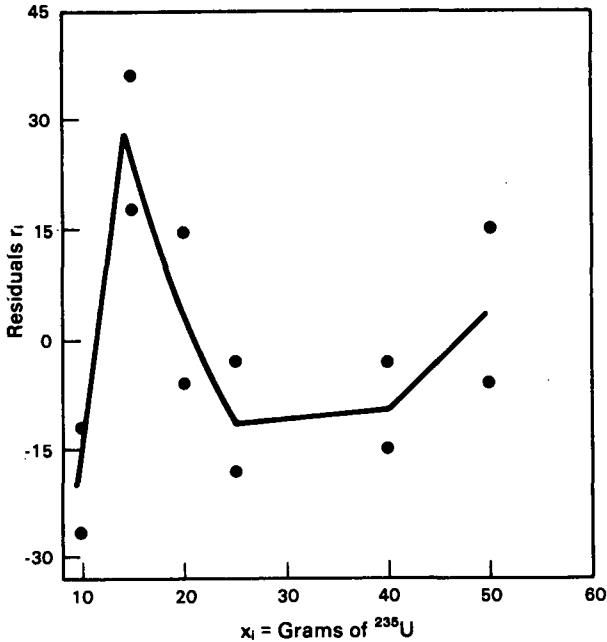


Figure 6.8 Scatterplot of residuals vs. the independent variable for the fitted quadratic model.

x_i , this conclusion is very subjective. A histogram and normal probability plot of the residuals from the fitted quadratic equation are shown in Figures 6.9 and 6.10, respectively. For this example, there are insufficient data (i.e., the sample size is too small) to reach any firm conclusions about normality from these two figures. However, the test for normality—presented by Filliben (1975) and based on the correlation coefficient of the normal probability plot—was performed, and the result did not contradict the assumption of normally distributed errors.

Middle of interval	Number of observations
-25	2 **
-5	8 *****
15	3 ***
35	1 *

Figure 6.9 Histogram of quadratic regression equation residuals.

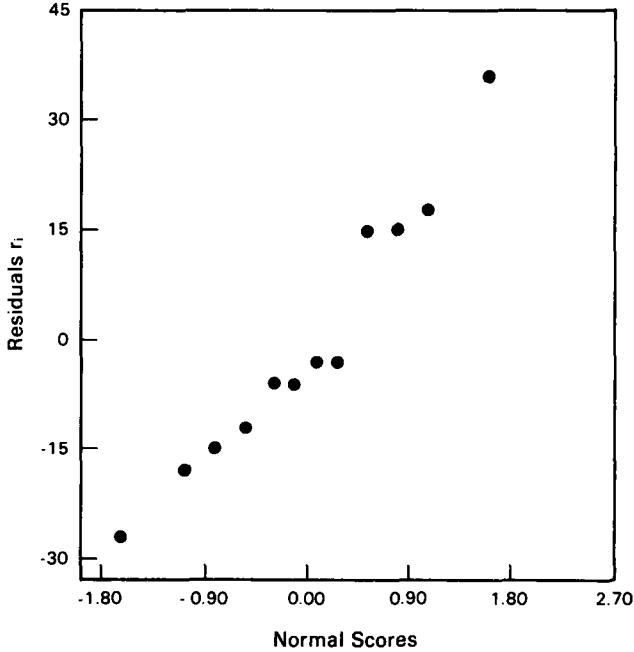


Figure 6.10 Normal probability plot of quadratic regression equation residuals.

Finally, the assumption of independent errors should be tested. (See Section 6.4.5.3 and Example 6.8).

6.4.5.2 Standardized Residuals

While the previously discussed residual plots help identify extreme residuals, they do not give all the information required for some situations. It can be helpful to consider standardized residuals, obtained by dividing each residual by its estimated standard deviation. The resulting standardized residual is given by

$$r_i^* = \frac{r_i}{s\sqrt{1 - h_{ii}}} \quad (6.32)$$

where

- r_i = the residual from the i^{th} data point
- s = the square root of the residual mean square of the regression, assuming the fitted equation has no significant lack of fit

The h_{ii} , $i = 1, 2, \dots, n$, are the diagonal elements of the so-called hat matrix $\underline{H} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$ (see Hoaglin and Welsch, 1978) and are given individually by

$$h_{ii} = \underline{x}_i'(\underline{X}'\underline{X})^{-1}\underline{x}_i \quad (6.33)$$

where \underline{x}_i' is the i^{th} row of \underline{X} . For a simple linear regression model, Equation 6.33 simplifies to the form

$$h_{ii} = [1, x_i] (\underline{X}'\underline{X})^{-1} \begin{bmatrix} 1 \\ x_i \end{bmatrix} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

The interested reader can work through the algebra by noting the form of the $(\underline{X}'\underline{X})^{-1}$ matrix in Equation 6.22. The standardized residuals can be plotted in the same manner as any of the residual plots discussed in the previous subsection.

In addition to graphical displays of the standardized residuals, approximate hypothesis tests can be applied. If the fitted regression equation does not have a significant lack of fit and the assumption of normally distributed errors appears reasonable, standardized residuals can be compared with $t_{1-\alpha/2}(n-p)$, obtained from Table A5, to determine whether an r_i^* is farther away from zero than would be expected by chance alone. Any observation corresponding to a significantly large r_i^* should be investigated to see whether there is an explanation for the large value. Note that, by chance alone, approximately 100 $\alpha\%$ of the standardized residuals are expected to be greater than $t_{1-\alpha/2}(n-p)$ or less than $-t_{1-\alpha/2}(n-p)$.

Lund (1975) presents an outlier test for the largest standardized residual. If r_i^* exceeds the critical table value, the corresponding observation is determined to be an outlier. Critical values for Lund's test are given in Table A9.

Example 6.7 Investigate the standardized residuals from the linear and quadratic fitted equations in Example 6.6.

The standardized residuals for both the linear and quadratic equations were obtained from the output of a statistical computer software package and are displayed in Table 6.6. The values for the quadratic equation can also be calculated by using the intermediate computer results tabulated in Example 6.9. Using an approximate t-test (the t-distribution critical values

TABLE 6.6
Standardized Residuals from the Fitted
Linear and Quadratic Equations

x_i	y_i	Standardized residuals from linear equation	Standardized residuals from quadratic equation
10	890	0.89	-1.66
10	905	1.13	-0.75
15	1234	0.73	1.97
15	1216	0.45	1.02
20	1491	-0.73	-0.28
20	1511	-0.43	0.77
25	1815	-1.15	-0.08
25	1798	-1.41	-1.02
40	2896	-0.80	-0.19
40	2886	-0.96	-0.74
50	3718	1.18	-0.35
50	3740	1.56	1.07

$t_{0.975}(10) = 2.23$	$t_{0.975}(9) = 2.26$
------------------------	-----------------------

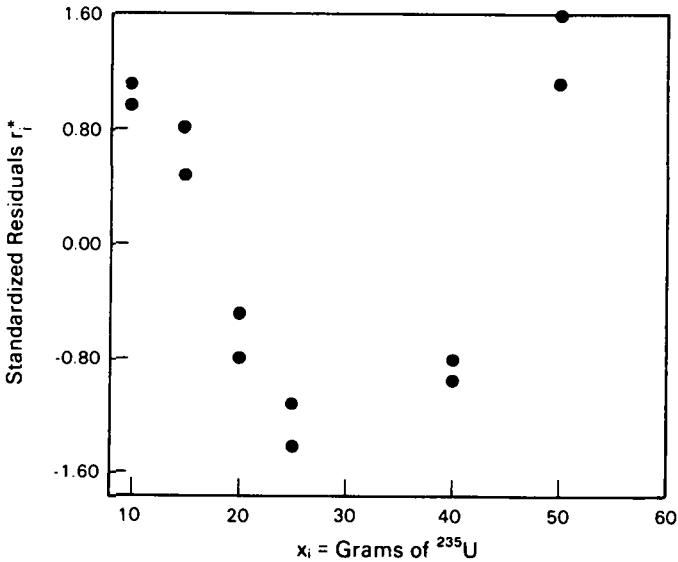


Figure 6.11 Standardized residuals from the fitted linear equation vs. the independent variable.

are given at the bottom of Table 6.6), none of the standardized residuals for either equation appears to be significantly different from zero at the $\alpha = 0.05$ level of significance. Because the critical values of Lund's

procedure, given in Table A9, are larger than all of the standardized residuals in Table 6.6, none of the residuals is classified as an outlier.

Plots of the standardized residuals against the independent variable x for both the fitted linear and quadratic equations are displayed in Figures 6.11 and 6.12, and are very similar in appearance to Figures 6.7 and 6.8. In this example, examination of the standardized residuals conveys essentially the same information as the unstandardized residuals. With the standardized residuals, however, it is easier to see their relative sizes and relationship to zero at a glance.

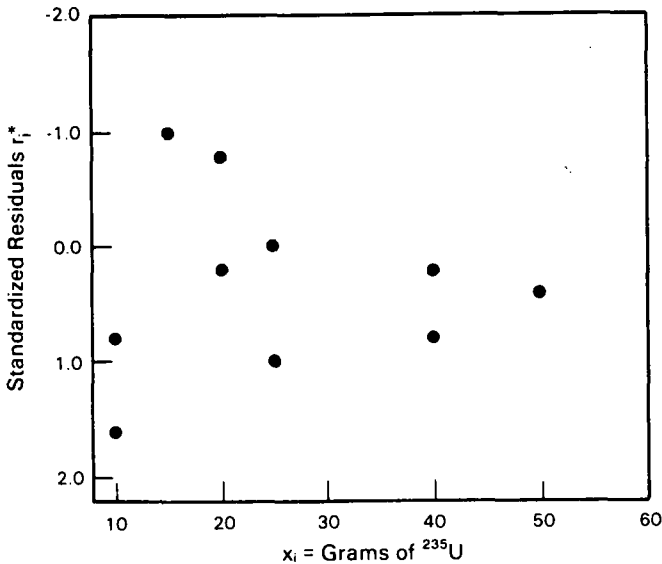


Figure 6.12 Standardized residuals from the fitted quadratic equation vs. the independent variable.

6.4.5.3 Serial Correlation

In applications of regression, the observational errors are often assumed to be independent, or at least pairwise uncorrelated. Situations where this assumption is not valid are often detected by plotting the residuals against the observation order. However, it is also desirable to apply statistical tests that detect departures from this assumption. Two such tests are considered in this section; they are the runs test and the Durbin-Watson test.

If observational errors are not independent, there are many ways that they might be related. They might be **serially correlated**; that is, the correlations between errors which are s steps apart (ρ_s) are always the same.

Assuming that the fitted regression equation under examination has no significant lack of fit, the residuals can be thought of as the observed errors from the correct model (i.e., a sample of the observational errors). In this situation, it is reasonable to investigate the residuals for serial correlation.

The correlation between residuals one (or two, three, ...) step(s) apart is called the lag-1 (or lag-2, lag-3, ...) serial correlation. The lag- s serial correlation is displayed by plotting each residual except the first s of them against the residual s steps preceding it. Calculating the sample correlation coefficient (see Equation 6.2, Section 6.1.1) for these pairs of residuals yields the lag- s serial correlation coefficient $\hat{\rho}_s$.

It is possible to obtain both positive and negative serial correlations. A positive serial correlation results when successive residuals tend to be alike. This is common, for example, when a measurement instrument drifts in one direction; that is, it becomes progressively more biased over time. Negative serial correlation results when successive residuals are different. This occurs, for example, when a system is subject to holdup (i.e., in successive measurement periods material alternately builds up and then is recovered).

The runs test. When the observation order of a set of residuals is known, it is sometimes evident that groups of positive or negative residuals occur in unusual patterns. If the underlying error structure really consists of independent errors, a random pattern of positive and negative residuals would be expected when the residuals are plotted in observational order. Thus, a certain number of runs of positive and negative residuals would be expected if indeed the assumption of independence is true. A **run** of positive (or negative) residuals is a sequence of successive residuals that are all positive (or negative). Fewer runs than the expected number would indicate the possibility of positive serial correlation among the observational errors, while more runs than the expected number would indicate the possibility of negative serial correlation.

It is of interest to test the hypothesis

$$H_0: \text{Errors are independent}$$

against the alternative

$$H_A: \text{Errors are not independent}$$

Let

n_1 = the number of positive residuals

n_2 = the number of negative residuals

n = $n_1 + n_2$ = the total number of residuals

U^* = the number of runs of positive and negative residuals

Given n_1 and n_2 , H_0 is rejected if the number of runs U^* is too large or too small. The details of the runs test are given in Section 9.1.

The Durbin-Watson test. The Durbin-Watson test is designed to detect a certain type of serial correlation. The hypotheses for this test are stated as

$$\left. \begin{array}{l} H_0 : \rho_s = 0 \\ H_A : \rho_s = \rho^s \end{array} \right\} s = 1, 2, \dots, \text{ for some } \rho \neq 0, -1 \leq \rho \leq 1 \quad (6.34)$$

The alternative hypothesis H_A specifies the presence of serial correlation which is either $+1$ or -1 (if $\rho = +1$ or -1) or which decreases in absolute value with increasing lag (if $-1 < \rho < 1$ and $\rho \neq 0$). The two-sided test of H_0 against H_A consists of computing the test statistic

$$D^* = \frac{\sum_{i=2}^n (r_i - r_{i-1})^2}{\sum_{i=1}^n r_i^2} \quad (6.35)$$

and comparing it to lower and upper critical values (d_L , d_U) found in Table A10 for significance levels of 0.05, 0.025, or 0.01. The critical values are given for various numbers of observations n , and for $k = 1, 2, 3, 4, 5$ predictor variables x_j . It is assumed that the residuals r_i , given by Equation 6.31, are obtained from fitting a model of the form

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i \quad (6.36)$$

by the least squares method to a set of experimental or sampling data, where x_j , $j = 1, 2, \dots, k$, are the predictor variables. Many simple and multiple regression models, including polynomial models, can be written in this form.

The two-sided Durbin-Watson test for serial correlation is as follows:

$$\left. \begin{array}{l} \text{If } D^* < d_L \text{ or } D^* > 4 - d_L, \text{ reject } H_0 \text{ at significance level } 2\alpha. \\ \text{If } D^* > d_U \text{ and } D^* < 4 - d_U, \text{ do not reject } H_0 \text{ at significance} \\ \text{level } 2\alpha. \\ \text{Otherwise, the test is inconclusive.} \end{array} \right\} \quad (6.37)$$

One-sided tests ($\rho > 0$ or $\rho < 0$) are discussed in Draper and Smith (1981) and Durbin and Watson (1951).

The possibility of obtaining an inconclusive result from the Durbin-Watson test is not attractive. Based on an approximate test from Durbin and Watson (1971), Draper and Smith (1981) suggest that an inconclusive result be replaced with rejection of H_0 , with the understanding that the actual probability of a Type I error is somewhat larger than the originally specified level.

Example 6.8 Continuing Example 6.6, assume that the observation order of the 12 observations is as given in Table 6.7, and that the residuals r_i are from the fitted quadratic regression equation. Determine whether the assumption of independent errors appears reasonable.

TABLE 6.7
Observation Order and Number
of Runs for the Fitted
Quadratic Equation

Order	x_i	y_i	r_i	Number of runs
1	15	1216	19.3	1
2	50	3718	-5.5	
3	10	905	-12.3	
4	25	1815	-1.6	
5	40	2896	-3.5	2
6	20	1511	14.6	3
7	25	1798	-18.6	4
8	50	3740	16.5	5
9	10	890	-27.3	
10	40	2886	-13.5	6
11	15	1234	37.3	7
12	20	1491	-5.4	8

First, consider the observation sequence plot of the residuals displayed in Figure 6.13. No time related effect is apparent.

Now consider the runs test for serial correlation in the residuals. There are $U^* = 8$ runs in the signs of the residuals as shown in Table 6.7, with $n_1 = 4$ positive residuals and $n_2 = 8$ negative residuals. The critical values, given in Table A11, are used to test the hypothesis

H_0 : number of runs observed is as expected

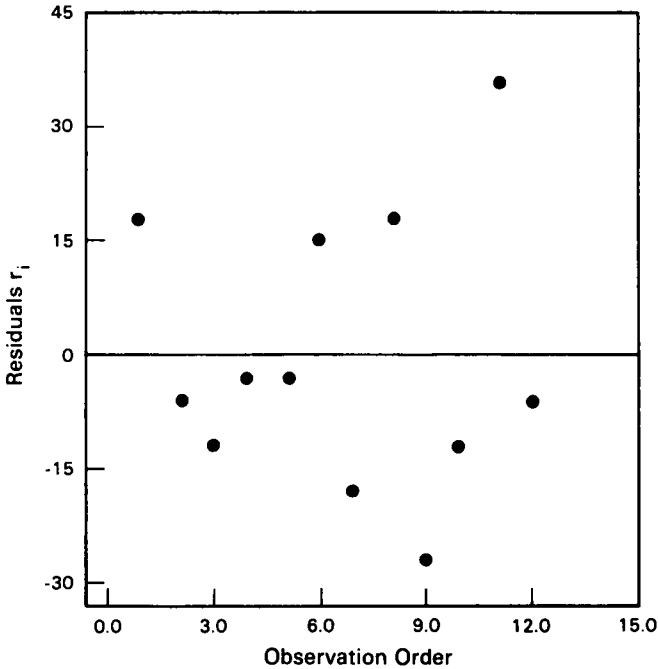


Figure 6.13 Observation sequence plot of residuals for the fitted quadratic equation.

against the alternative

H_A : too few or too many runs are observed

Note that when H_0 is true, the probability of observing 8 or more runs is given by

$$\Pr[U \geq 8] = 1 - \Pr[U \leq 7] = 1 - 0.788 = 0.212$$

and the probability of observing 8 or fewer runs is given by

$$\Pr[U \leq 8] = 0.929$$

Thus, observing 8 runs in the signs with $n_1 = 4$ and $n_2 = 8$ is not unusual enough to reject H_0 (if $\alpha \leq 0.2$). It is concluded that the residuals (and hence the errors) do not appear to be serially correlated according to the runs test.

For illustration only, the Durbin-Watson test statistic is computed by applying Equation 6.35 which gives

$$D^* = \frac{\sum_{i=2}^n (r_i - r_{i-1})^2}{\sum_{i=1}^n r_i^2} = \frac{9954.05}{3748.20} = 2.66$$

However, because Table A10 begins with $n = 15$ and we have $n = 12$, the Durbin-Watson test is not applicable to this example.

To conclude this example, the time order plot and the runs test seem to indicate that the assumption of independent errors is reasonable.

When significant serial correlation is found by the runs test or the Durbin-Watson test, the basic method of analysis is weighted least squares (Section 6.3.2). However, this requires obtaining the weight matrix \underline{V} (Equation 6.17). Draper and Smith (1981) provide a method for developing an estimate of \underline{V} .

6.4.6 Detection of Outliers and High Leverage Observations

When fitting a regression equation to data, an outlying observation can significantly affect parameter estimates obtained by least squares estimation. If the observed y value is quite different from what would normally be observed under similar conditions (e.g., the difference may be due to an instrument failure or transposition of numbers in recording the observation) and if it significantly affects the least squares coefficient estimates, then an erroneous fitted equation can result. Such an outlying observation should be detected, its effect on the coefficient estimates determined, its cause determined and corrected if possible, or a decision should be made about whether to retain it as part of the data.

It is also helpful to detect data points that may influence the regression analysis due to the values of the independent variables. A point with independent variable values considerably removed from other values in the data set is said to have high leverage and may be quite influential in fitting an equation to the data, especially if the observed y value is unusual.

As an illustration of the concepts of high leverage and outlying observations, consider the four (x,y) scatterplots in Figure 6.14. In (a), there are no high leverage or outlying observations. In (b), there is an outlying observation that does not have high leverage; while in (c), there is a high

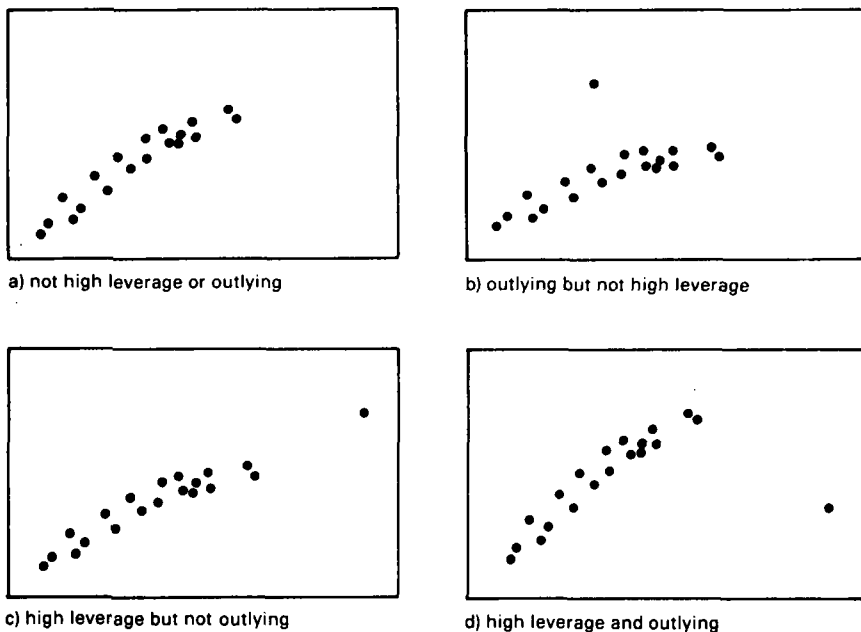


Figure 6.14 Examples of high leverage and outlying observations.

leverage observation that is not outlying. Finally, in (d), there is an observation that is outlying and has high leverage.

There are many techniques for detecting outlying or influential data points. It is sometimes possible to detect outlying observations by viewing a plot of y vs. the independent variables or any of the previously discussed residual plots. However, this becomes very difficult when there are several independent variables; quantitative techniques that are more objective are required in these situations. Whichever technique is used, outlying or influential observations should not be removed from the data set unless traceable to an assignable cause, such as a faulty instrument or a transposition or keypunch error. Some “outlying” observations are actually correct and provide valuable information.

6.4.6.1 Detecting High Leverage Points

In this subsection, a measure of the leverage exerted on \hat{y}_i by y_i is presented. This leverage depends only upon the values of the independent variables. Substantial leverage implies that y_i (regardless of its observed value) has a significant impact on \hat{y}_i and possibly on the coefficient estimates in the fitted regression equation. Although leverage of a data point is discussed in terms of the impact of the observed y_i , this impact depends on the x_{ij} values. A point has high leverage only when one or more of the independent variables have extreme values.

The measure of leverage for the i^{th} data point is given by h_{ii} , the i^{th} diagonal element of the hat matrix given in Equation 6.33. Hoaglin and Welsch (1978) suggest that a value of h_{ii} is large (note that $0 \leq h_{ii} \leq 1$) if

$$h_{ii} > \frac{2p}{n} \quad (6.38)$$

where n is the number of data points and p is the number of coefficients in the regression model. Because the hat matrix does not involve the y_i , this test for points with high leverage involves only the x values of each data point. High leverage indicates that a point has the potential for unduly influencing the parameter estimates. However, the observed y value is considered in determining whether the point is actually influential.

6.4.6.2 Measures of Influence

The effect of any data point on a fitted regression equation (its influence) can be measured by deleting the point from the data and refitting the equation. One measure of influence based on this concept, presented by Cook (1977), proposes that the measure of influence of the i^{th} data point be the distance

$$D_i = \{[\hat{\beta} - \hat{\beta}_{(i)}]'X'X[\hat{\beta} - \hat{\beta}_{(i)}]\}/(ps^2) \quad (6.39)$$

where

- X = the $n \times p$ matrix of predictor variables
- $\hat{\beta}$ = the $p \times 1$ vector of least squares parameter estimates obtained by fitting the equation to all data
- $\hat{\beta}_{(i)}$ = the $p \times 1$ vector of least squares parameter estimates obtained by fitting the equation to the remaining data after omitting the i^{th} data point
- s^2 = the residual mean square from the fitted equation based on all data points
- p = the number of unknown coefficients in the regression equation

A reduced form of Equation 6.39, suitable for computation, is given by

$$D_i = \frac{(r_i^*)^2}{p} \left(\frac{h_{ii}}{1 - h_{ii}} \right) \quad (6.40)$$

where r_i^* is the i^{th} standardized residual given by Equation 6.32 and h_{ii} is given by Equation 6.33. A large D_i value compared to the remaining D_j ($j \neq i$) values indicates that the i^{th} observation is influential. Note that D_i can be large if either r_i^* is large or if $h_{ii}/(1 - h_{ii})$ is large; a large r_i^* indicates an unusual y_i value, while a large $h_{ii}/(1 - h_{ii})$ indicates a data point with high leverage values of the independent variables. A further discussion of D_i can be found in Cook (1977) or Cook (1979). Alternative approaches to evaluating influence are given by Belsley, Kuh, and Welsch (1980).

A point that has high leverage and whose y value is an outlier significantly affects the least squares parameter estimates. If an assignable cause is identified, the questionable point should either be corrected or deleted from the data set, after which the equation should be refitted. Even when an assignable cause (such as faulty equipment) is not identified, it may be enlightening to perform the regression analysis with the point omitted so that the results with and without the questionable point can be compared.

Example 6.9 Continuing the analysis of the quadratic regression equation obtained in Example 6.6, investigate the possibility of high leverage points or influential observations.

The h_{ii} , r_i^* , and D_i values for each of the 12 data points are displayed in Table 6.8.

TABLE 6.8
Data with Influence and Leverage Measures
for Fitted Quadratic Equation

i	x_i	y_i	\hat{y}_i	r_i	h_{ii}	r_i^*	D_i
1	10	890	917.3	-27.3	0.356	-1.66	0.51
2	10	905	917.3	-12.3	0.356	-0.75	0.10
3	15	1234	1196.7	37.3	0.143	1.97	0.22
4	15	1216	1196.7	19.3	0.143	1.02	0.06
5	20	1491	1496.4	-5.4	0.144	-0.28	0.00
6	20	1511	1496.4	14.6	0.144	0.77	0.03
7	25	1815	1816.6	-1.6	0.212	-0.08	0.00
8	25	1798	1816.6	-18.6	0.212	-1.02	0.09
9	40	2896	2899.5	-3.5	0.210	-0.19	0.00
10	40	2886	2899.5	-13.5	0.210	-0.74	0.05
11	50	3718	3723.5	-5.5	0.436	-0.36	0.03
12	50	3740	3723.5	16.5	0.436	1.07	0.29

By using the comparison criterion of h_{ii} given by Equation 6.38, with $n = 12$ and $p = 3$, there are no h_{ii} values larger than $2p/n = 0.50$.

Thus, there are no points where the x_i or x_i^2 values have extremely high leverage.

Cook's distance (D_i) is extreme for $i = 1$ compared to the remaining data points; thus, the first data point is deemed influential. Because it is not a high leverage point, however, it appears to be influential as a result of an unusual y value and should be investigated further.

6.4.7 Model Validation

So far, several techniques have been discussed for evaluating a fitted regression equation. All of these techniques are based on statistics calculated from the data set used to fit the model. Because a major use of regression in nuclear material control applications is for making predictions, the final test of the fitted equation's predictive ability should be based on a data set that was not used to fit the equation. Intuitively, this model validation process is very simple: a number of data points representative of the range of the independent and dependent variables should be collected and the observed y values compared with the values predicted by the fitted regression equation. If the predicted y values are satisfactorily close to the observed values, then the model is deemed valid.

Sometimes it is not practical or possible to obtain additional data for model validation. In such cases, prior to model fitting, the complete data set is split into two subsets by some reasonable criterion. One subset is used to carry out the regression analysis and model development process, as discussed in the previous sections of this chapter. Once a satisfactory prediction equation has been developed, the other data subset is used to validate it; that is, to see how well it predicts. There is no unique or best approach to this problem. However, some possibilities are briefly introduced below.

The PRESS procedure, introduced by Allen (1971), can be used for model validation. This procedure involves leaving out one observation, fitting the model to the rest of the data, predicting the y value for the omitted observation, and computing the residual $y - \hat{y}$. The omitted observation is then returned to the data set, and the process is repeated until each observation has been left out. The sum of squared residuals is then computed. This statistic is a valuable aid when choosing from among several candidate models; that is, a "best" model that has been validated over the data set can often be identified. Also, the individual residuals can be examined for inconsistencies.

Geisser (1975) considers a procedure similar to PRESS, but he extends the concept to leaving out m observations and using the $n - m$ remaining observations to fit the model. The m omitted observations are then used to validate the fitted model.

The practice of using half the data to develop a prediction model and the other half to validate it has been popular for many years. The DUPLEX algorithm, discussed by Snee (1977), splits the data according to a criterion which guarantees that the properties of the determinant of $\underline{X}'\underline{X}$ are similar for both the modeling half and the validation half of the data. Snee recommends that a data set should not be split in half unless the total sample size n is larger than $2p + 25$, where p is the number of unknown coefficients in the model to be fitted.

6.5 USING A FITTED EQUATION

While the data collection, parameter estimation, and model evaluation are all important components of a regression analysis, using the fitted equation is often the ultimate goal. The primary uses of a fitted equation are for prediction and interpretation of an underlying relationship. Prediction is of primary importance in nuclear material accounting. Methods for interpretation (response surface analysis) are presented by Myers (1976).

Using a fitted equation to predict y values at specified values of the predictor variables is quite easy. A predicted value is obtained from

$$\hat{y}_k = \hat{\beta}' \underline{x}_k = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_{p-1} x_{p-1} \quad (6.41)$$

where \hat{y}_k is the predicted value, $\hat{\beta}$ the vector of coefficient estimates, and \underline{x}_k is the vector of specific values of the predictor variables.

A predicted y value from a fitted equation is an observed value of a random variable due to errors from several sources (see Section 6.5.2), and therefore it has a sampling distribution with a mean and variance. The variance of a predicted value is dependent upon how the prediction is to be interpreted. The same predicted value \hat{y}_k could be interpreted either as a prediction of a single observation or as the mean of many observations taken at a particular set of independent variables. The variance of a mean is smaller than the variance of a single observation. The estimated variance of y_k , when interpreted as a mean prediction at \underline{x}_k , is given by

$$\hat{\text{Var}}_{\text{Fit}}[\hat{y}_k] = s^2[\underline{x}_k'(\underline{X}'\underline{X})^{-1}\underline{x}_k] \quad (6.42)$$

When \hat{y}_k is interpreted as the prediction of a single observation, the estimated variance is given by

$$\hat{\text{Var}}_{\text{Pred}}[\hat{y}_k] = \hat{\text{Var}}_{\text{Fit}}[\hat{y}_k] + s^2 \quad (6.43)$$

In Equations 6.42 and 6.43,

- \underline{x}_k = the $p \times 1$ vector of known predictor variables for which a prediction \hat{y}_k is desired
- \underline{X} = the $n \times p$ matrix of predictor variable values for the data set used to fit the equation
- s^2 = the residual mean square from the fitting the regression equation under investigation

Because simple linear regression is of particular interest in nuclear material control applications, the simplified forms of Equations 6.42 and 6.43 for a simple linear regression ($\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k$) are given by

$$\text{Vâr}_{\text{Fit}}[\hat{y}_k] = s^2 \left[\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \right] \quad (6.44)$$

and

$$\text{Vâr}_{\text{Pred}}[\hat{y}_k] = \text{Vâr}_{\text{Fit}}[\hat{y}_k] + s^2 \quad (6.45)$$

Whether \hat{y}_k is interpreted as the prediction of a mean or of a single value, a $100(1 - \alpha)\%$ confidence interval for the true value is given by

$$\hat{y}_k \pm t_{1-\alpha/2}(n-p) \sqrt{\text{Vâr}(\hat{y}_k)} \quad (6.46)$$

where $\text{Vâr}[\hat{y}_k]$ is given by either Equation 6.42 or 6.43, depending upon the interpretation of \hat{y}_k , and $t_{1-\alpha/2}(n-p)$ is obtained from Table A5. Note that the confidence interval given by Equation 6.46 is for the unknown true value of y_k corresponding to a single vector \underline{x}_k . To construct simultaneous $100(1-\alpha)\%$ confidence intervals for all possible sets of independent variable values, the appropriate formula is

$$\hat{y}_k \pm \sqrt{\text{Vâr}[\hat{y}_k] p F_{1-\alpha}(p, n-p)} \quad (6.47)$$

where $F_{1-\alpha}(p, n-p)$ is obtained from Table A6.

Example 6.10 Using the fitted quadratic equation from Example 6.6, predict the y_k value for $x_k = 19$ and estimate its variance. Then construct 95% confidence intervals for the true value of y_k , assuming first that y_k is a predicted mean and then assuming that y_k is a predicted individual value.

The predicted value of y_k at $x_k = 19$ is computed as

$$\begin{aligned}\hat{y}_k &= \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 \\ &= 419.83 + 45.665(19) + 0.408(19^2) = 1434.75\end{aligned}$$

If this \hat{y}_k is interpreted as a mean prediction when $x_k = 19$, then Equation 6.42 is applied to compute $\text{V\hat{a}r}_{\text{Fit}}(\hat{y}_k) = 56.10$. If \hat{y}_k is interpreted as the prediction of a single observation at $x_k = 19$, then $\text{V\hat{a}r}_{\text{Pred}}(\hat{y}_k) = 473.06$ is obtained by applying Equation 6.43.

Equation 6.46 yields 95% confidence intervals where $\hat{y}_k = 1434.75$, $\alpha = 0.05$, and $t_{0.975}(9) = 2.26$:

\hat{y}_k as the prediction of a mean	\hat{y}_k as the prediction of a single observation
$1434.75 \pm 2.26 (7.49)$	$1434.75 \pm 2.26 (21.75)$
1434.75 ± 16.93	1434.75 ± 49.15
$(1417.82, 1451.68)$	$(1385.60, 1483.90)$

Note that if simultaneous 95% confidence intervals were of interest, application of Equation 6.47 yields:

\hat{y}_k as the prediction of a mean	\hat{y}_k as the prediction of a single observation
$1434.75 \pm \sqrt{56.10(3)(3.86)}$	$1434.75 \pm \sqrt{473.06(3)(3.86)}$
1434.75 ± 25.49	1434.75 ± 74.01
$(1409.26, 1460.24)$	$(1360.74, 1508.76)$

REFERENCES

- Allen, D. M., 1971, *The Prediction Sum of Squares as a Criterion for Selecting Predictor Variables*, Technical Report No. 23, Department of Statistics, University of Kentucky, Lexington, Kentucky.

- Anscombe, F. J., 1961, Examination of Residuals, *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, 1: 1-36.
- , and J. W. Tukey, 1963, The Examination and Analysis of Residuals, *Technometrics*, 5: 141-160.
- Belsley, D. A., E. Kuh, and R. E. Welsch, 1980, *Regression Diagnostics*, John Wiley & Sons, Inc., New York.
- Berkson, J., 1950, Are There Two Regressions?, *J. Am. Stat. Assoc.*, 45: 164-180.
- Cook, R. D., 1977, Detection of Influential Observations in Linear Regression, *Technometrics*, 19: 15-18.
- , 1979, Influential Observations in Linear Regression, *J. Am. Stat. Assoc.*, 74: 169-174.
- , and S. Weisberg, 1982, *Residuals and Influence in Regression*, Chapman and Hall, New York.
- Daniel, C., and F. S. Wood, 1980, *Fitting Equations to Data*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Draper, N. R., and H. Smith, 1981, *Applied Regression Analysis*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Durbin, J., and G. S. Watson, 1951, Testing for Serial Correlation in Least Squares Regression II, *Biometrika*, 38: 159-178.
- , and G. S. Watson, 1971, Testing for Serial Correlation in Least Squares Regression III, *Biometrika*, 58: 1-19.
- Filliben, J. J., 1975, The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*, 17: 111-117.
- Geisser, S., 1975, The Predictive Sample Reuse Method with Applications, *J. Am. Stat. Assoc.*, 70: 320-328.
- Graybill, F. A., 1969, *Introduction to Matrices with Applications in Statistics*, Wadsworth, Belmont, California.
- , 1976, *Theory and Application of the Linear Model*, Duxbury Press, Boston.
- Hoaglin, D. C., and R. E. Welsch, 1978, The Hat Matrix in Regression and ANOVA, *Am. Stat.*, 32: 17-22.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Lund, R. E., 1975, Tables for an Approximate Test for Outliers in Linear Models, *Technometrics*, 17: 473-476.
- Myers, R. H., 1976, *Response Surface Methodology*, Virginia Polytechnic Institute, Blacksburg, Virginia.
- Searle, S. R., 1966, *Matrix Algebra for the Biological Sciences*, John Wiley & Sons, Inc., New York.
- , and W. H. Hausman, 1970, *Matrix Algebra for Business and Economics*, John Wiley & Sons, Inc., New York.
- Seber, G. A. F., 1977, *Linear Regression Analysis*, John Wiley & Sons, Inc., New York.
- Snee, R. D., 1977, Validation of Regression Models: Method and Examples, *Technometrics*, 19: 415-428.

CHAPTER 7

Experimental Design

7.0 INTRODUCTION

The subject matter in most chapters of this book is oriented toward applications of statistical estimation and hypothesis testing. Because hypotheses are tested and parameters are estimated using data from experiments, the manner in which an experiment is performed is extremely important. An **experiment** is an action or process used to collect data. The plan which specifies the method and the order in which an experiment is performed is referred to as the **experimental design**.

The **statistical design of experiments** is a process by which data collection is planned, organized, and conducted so that the questions motivating the data collection can be answered by a statistical analysis of the data. This definition indicates that there are three basic aspects to an experimental problem: the design or planning of the experiment, the collection of data during the experiment, and the statistical analysis of the data (taking into account any departures from the original design). These subjects are closely related because the method of analysis depends directly upon the design actually employed and the data collected. It is often impossible, however, to salvage valid conclusions from an undesigned or poorly designed experiment, no matter which statistical analysis methods are applied.

The principles of experimental design have many applications in nuclear material control problems. Two major areas of application are calibration work and the investigation (and estimation) of the extent to which specified factors contribute to a total measurement uncertainty. As an illustration of the second area of application, consider a situation where sampling variability and analytical variability both contribute to the total measurement uncertainty in determining the percent uranium factor for a can of UO_2 powder. Experimental design principles can be applied so that data are collected in a manner which accommodates separate estimation of the contributions of sampling and analytical variability to the total measurement variability.

Analysis of variance methods, including those presented in Chapter 5, are applicable to the analysis of data from the experimental designs presented in this chapter. Thus, computational formulas and specific details of hypothesis testing and parameter estimation are not given in this

chapter for each design. Instead, references are made to specific sections of Chapter 5 and to other texts where the reader can find appropriate computational formulas and statistical methods for analyzing data from a given design. Computer software packages are available for analysis of data from designed experiments. Although such packages are highly recommended, care must be exercised in choosing one that is well suited to the application at hand.

Section 7.1 presents the major steps involved in designing an experiment. In the remaining sections of this chapter, some commonly used experimental designs are presented. It should be noted that the experimental design concepts and plans presented in this chapter are necessarily limited because an entire book would be required for a more complete presentation. The intent of this chapter is to introduce a few basic principles of experimental design and to briefly describe several specific designs. For some of the less sophisticated designs, enough information is presented for the reader to use them without additional reference material. Some of the more sophisticated designs are only briefly introduced, however, with the expectation that a reader should refer to the given references to utilize the designs. Thus, this chapter gives an introduction to experimental design, an appreciation for its importance, and references that provide detailed discussions of individual topics, including guidance for more complicated applications than those presented here.

7.1 EXPERIMENTAL DESIGN BASICS

The structure of an experimental design is defined in terms of **replication**, **randomization**, **experimental units**, and **isolation**. Each of these is considered below.

Replication is the repetition of all or part of an experiment. For example, in the instrument calibration problem described in Example 6.5, each quantity of ^{235}U is measured twice by the NDA instrument. If the experiment is defined as the measurement of the six ^{235}U standards, then the experiment is replicated twice. Replication has many advantages. It allows the experimenter to estimate the experimental error variance, which is an important quantity in the analysis of variance methods as applied to experimental data (Chapter 5). Replication improves the precision of results obtained from the statistical analysis of the experimental data. It is also helpful in detecting gross errors in the data.

Randomization refers to the random determination of the order in which individual **runs** or **trials** of an experiment are to be performed and to the random allocation of **experimental units** to the individual runs or trials. **Experimental units** are items or individual portions of material to be tested or measured during an experiment. The investigation or measurement of a single experimental unit is referred to as a **run** or **trial**.

Randomization supports the assumption of most statistical analysis techniques that the observations (or errors) are independently distributed random variables. Randomizing an experiment also helps to evenly distribute the effects of factors not studied or controlled in the experiment, so that they can be treated as part of the random experimental error. This reduces the chance of an uncontrolled factor biasing a particular set of experimental results.

Isolation refers to the structuring of the experimental design so that the quantities of interest (e.g., the overall mean, main effects, and/or variance components, as defined in Chapter 5) can be separately estimated and tested. As an illustration, consider an experiment to compare two methods of sampling plutonium nitrate solution from a storage tank. It is believed that sampling variability and variability in the analytical method for measuring the plutonium concentration may affect the results. The experimental design must be structured so that the effects of sampling, the sampling method, and the analytical method can be isolated from each other and separately estimated.

Now that some of the building blocks of an experimental design have been introduced, consider the design of an experiment from a broader viewpoint. The design of a particular experiment can be summarized in the following major steps, all of which must be completed before the experiment is conducted:

1. Definition of the specific experimental objectives.
2. Determination of the required number of observations that must be taken to achieve either a specified power of the analysis of variance tests or a specified level of precision of parameter estimates.
3. Definition of the experimental procedure.
4. Review of the results of Steps 2 and 3 with budget and facilities considerations to assess feasibility.
5. Identification of an appropriate statistical analysis procedure.

Each of these steps is discussed below.

Step 1. Definition of experimental objectives. First, a clear statement is needed of the problem being investigated. The statement of the problem should include the identification of the **dependent** or **response variables** to be studied and the **independent variables** or **factors** that may affect them. The specific objectives of the experiment should be clearly stated in terms of these response variables and factors.

Several questions should be answered to help define the experiment. Can the dependent variables be measured? If so, how accurately and precisely and with what kind of instrument? If not, what type of response is expected (yes/no, ranks, etc.)? How are the levels of independent variables (factors) to be selected? Is each factor to be held constant, to assume certain specified levels, or to assume levels chosen at random from

all possible levels? If a factor is to be varied over several levels, are the levels to be defined quantitatively or qualitatively? Are the levels of a factor appropriate and adequate for the isolation, estimation, and testing of its true effect on the response variable?

Step 2. Determination of the number of observations to be taken. Considerations affecting the number of observations to be taken include the size of differences to be detected (in comparative studies), the precision required of an estimate (in estimation studies), the amount of variation expected in experimental results, and the probabilities of Type I and Type II errors that will be tolerated in the analysis of variance tests. Often, however, the overriding considerations are time, money, and available facilities. This is discussed further under Step 4.

Step 3. Definition of the experimental procedure. The randomization procedures to be used in the experiment must be specified. Is the whole experiment to be completely randomized? Is the experiment to be performed in portions necessitating randomization among the portions and within each portion? Recall that randomization involves randomizing the order of individual trials of the experiment as well as randomly assigning experimental units to the trials.

Step 4. Assessment of feasibility. Designing an experiment is often an iterative process. In the initial phases of planning an experiment, Steps 1, 2, and 3 might be completed. Then the required costs, manpower, and facilities must be compared with the budget and available resources. If the requirements exceed what is available, then the scope of the experiment must be reduced or the objectives must be modified so that a smaller, less expensive experiment is adequate. Example 4.3 illustrates this idea.

Sometimes Steps 1, 2, 3, and 4 must be repeated several times before a final experimental plan is approved for implementation. When an experimental design is modified, however, it is important that all parties involved understand the properties and limitations of the associated hypothesis tests and parameter estimates. Power curves or OC curves are helpful in illustrating the effect of changing an experimental design.

Step 5. Identification of an appropriate statistical analysis procedure. An appropriate statistical analysis procedure for the proposed experimental design should be identified and reviewed to verify that it is possible to analyze the experimental data in such a way that all quantities of interest can be estimated and/or tested without restrictive qualifications and assumptions. If this cannot be verified, then the results from Steps 1 through 4 should be reviewed to determine how the experimental design, the objectives, or both must be modified to accommodate a meaningful, worthwhile analysis of the data.

This chapter presents several types of experimental designs. These designs are characterized mainly by the number of factors (independent

variables) and the randomization procedure used. Most of the designs are multipurpose; that is, they are useful for comparing levels of fixed effects factors, estimating variance components of random effects factors, and estimating parameters in regression equations. The designs in the following sections are presented so that they are accessible to a nuclear material accounting experimenter. The experimenter is expected to consider the definition of the problem and its objectives so that an applicable experimental design can be chosen from among those discussed in the following sections, or, if necessary, from the references given.

7.2 DESIGNS FOR SINGLE FACTOR EXPERIMENTS

In this section, designs are considered for experiments involving one dependent (response) variable and one independent variable (factor). These designs could be applied to a simple linear regression problem or to several types of analysis of variance (ANOVA) applications. For example, an experiment might consist of measuring a known standard repeatedly at each of several times to determine whether an instrument requires adjustment. For this experiment, the single factor is time. Another experiment might consist of selecting a sample of UO_2 powder from each of several cans. Each sample is split into two aliquots, and the aliquots are randomly assigned to two different laboratories for analysis. Here the analysis capabilities of the two laboratories are to be compared, where "laboratory" is the single factor under investigation. Note that another factor (can of UO_2) exists in this experiment. Factors of this type, whose variability is accounted for in the statistical analysis but is not of primary concern, are referred to as **blocking factors**. The use of blocking factors in single factor experimental designs is discussed in Sections 7.2.2 and 7.2.3.

7.2.1 Completely Randomized Designs

A **completely randomized design** for a single factor experiment imposes no restrictions on randomization of the experimental runs, or if applicable, on the experimental units or material assigned to each run. This design is appropriate when there are no blocking factors (the experimental units or material are homogeneous) and when no trends in the response over time are expected. If the experimental units are not homogeneous and one or more blocking factors can be identified, the designs discussed in Sections 7.2.2 or 7.2.3 provide increased power or sensitivity of hypothesis tests. Cochran and Cox (1957) point out that complete randomization may also be appropriate when an appreciable fraction of the units is likely to be destroyed or fails to respond, and in small experiments where the increased power of hypothesis tests from designs incorporating blocking factors does

not outweigh the loss of error degrees of freedom due to blocking; that is, the error mean square is not reduced.

The statistical model for a completely randomized single factor design is given by

$$\begin{aligned}
 y_{ij} &= \mu + \alpha_i + \epsilon_{ij} & (7.1) \\
 i &= 1, 2, \dots, a \\
 j &= 1, 2, \dots, n
 \end{aligned}$$

where

- y_{ij} = the j^{th} observation at the i^{th} level of the factor of interest
- μ = the overall mean
- α_i = the effect of the i^{th} level of the factor
- ϵ_{ij} = the random experimental error affecting y_{ij}

The development and statistical analysis of a completely randomized single factor design is illustrated in the following example. The computational formulas used in this example are presented in Section 5.1.1.

Example 7.1 Consider an experiment that consists of making non-destructive measurements for percent of plutonium on mixed oxide pellets. Assume that 21 pellets from a batch of homogeneous pellets are to be measured using three similar but distinct nondestructive assay (NDA) instruments (seven pellets by each instrument). The purpose of the experiment is to compare the three instruments by testing the hypothesis that they produce equivalent results. Set up a completely randomized design for this experiment, and test the hypothesis of interest at the 0.05 level of significance.

For a completely randomized design, the seven pellets measured by each instrument are assumed to be replicate observations. Further, it is assumed that the 21 pellets are selected at random from a batch of homogeneous pellets. Let y_{ij} represent the measured response from the trial of the experiment for the j^{th} replicate using the i^{th} instrument, where $i = 1, 2, 3$ and $j = 1, 2, \dots, 7$, as illustrated in Table 7.1. The order in which the 21 trials should be run is determined by randomly selecting integers between 1 and 21, one at a time and without replacement (procedures for doing so are discussed in Chapter 11), and assigning them to the trials. One possible outcome of this process is illustrated in Table 7.1, which shows that y_{23} would be measured first, y_{36} second, \dots , and y_{24} last. The measured percent Pu values for this experiment are given in Table 7.2.

TABLE 7.1
Summary of One Possible Completely
Randomized Experimental Design

Replicate	Instrument		
	1	2	3
Identification of the trials			
1	Y_{11}	Y_{21}	Y_{31}
2	Y_{12}	Y_{22}	Y_{32}
3	Y_{13}	Y_{23}	Y_{33}
4	Y_{14}	Y_{24}	Y_{34}
5	Y_{15}	Y_{25}	Y_{35}
6	Y_{16}	Y_{26}	Y_{36}
7	Y_{17}	Y_{27}	Y_{37}
One possible completely randomized run order*			
1	10	11	20
2	19	14	9
3	12	1	4
4	6	21	13
5	15	17	5
6	18	16	2
7	3	8	7

*Pellet identification numbers are given.

TABLE 7.2
Percent by Weight of Pu in
Mixed Oxide Pellets

Replicate	Instrument		
	1	2	3
1	2.51	2.53	2.50
2	2.46	2.50	2.46
3	2.47	2.50	2.47
4	2.49	2.53	2.53
5	2.51	2.49	2.53
6	2.45	2.42	2.50
7	2.47	2.45	2.45
\bar{y}_i	2.480	2.489	2.491

The analysis of variance results are displayed in Table 7.3. The calculation of the entries in the ANOVA table is discussed in Section 5.1.1 (Table 5.2). Table A6 does not give a value of $F_{0.95}(2,18)$; however, the values $F_{0.95}(2,15) = 3.68$ and $F_{0.95}(2,20) = 3.49$ are given and can be used with linear interpolation to compute

$$F_{0.95}(2,18) \cong 3.49 + (3.68 - 3.49) \frac{(20-18)}{(20-15)} = 3.57$$

Because $F = 0.23 < 3.57$, there is no indication that the instruments are different. In fact, $F = 0.23$ is somewhat low, which suggests that the experimental error variance may be inflated. This topic is considered further in Example 7.2 in the following section.

TABLE 7.3
One-Way Fixed Effects ANOVA Table
for a Single Factor Experiment

Source	df	SS	MS	F
Instruments	2	0.00050	0.00025	0.23
Error	18	0.01957	0.00109	
Total	20	0.02007		

It should be noted that, because $F = 0.23$ is less than both $F_{0.95}(2,15)$ and $F_{0.95}(2,20)$, the interpolation is not necessary to determine the outcome of the hypothesis test. The interpolation was carried out, however, for illustration.

Although simple to set up, the principal objection to a completely randomized design is that it can result in less powerful ANOVA tests and wider confidence intervals for means than other designs. Because the randomization is not restricted in any way, all of the variation among experimental units enters into MS_{Error} , the estimate of the experimental error variance. If the variability among experimental units is quite large, the effects of the factor being investigated might not be detected. This idea was discussed in connection with the paired t-test in Section 4.4.

7.2.2 Randomized Complete Block Designs

In many experimental problems, it is advisable to isolate variability arising from extraneous sources through the use of blocking factors. A **randomized complete block design** isolates extraneous variability from one source (with one blocking factor). The essence of this design is that the experimental units are divided into homogeneous groups called **blocks**. In experimental design terminology, the term “complete” means that each block includes at least one observation at every level of the factor being investigated. **Randomized incomplete block designs** are discussed briefly in Section 7.2.4.

Grouping experimental units into homogeneous blocks tends to minimize experimental error variability within each block. Then, if the experimental units within a block are randomly assigned to the levels of the factor being investigated, the variability within blocks and the variability between blocks can be separately estimated. To illustrate this, suppose that in Example 7.1 the 21 pellets actually come from seven different batches (3 pellets from each batch). With a completely randomized design, the between-batch variability is included in the estimate of the experimental error variance, in addition to the variability among pellets within the batches. A randomized complete block design would allow the between-batch variability to be isolated and separately estimated so that it does not enter into the estimate of the experimental error variance.

The randomization of the experimental units in a randomized complete block design is carried out within the blocks. That is, the experimental units within a block are randomly assigned to the levels of the factor being investigated. At least one experimental unit within each block must be assigned to each factor level.

The principle of randomization is also applied in determining the run order of the experimental trials. The recommended approach is to randomize the order of trials within blocks, where all observations within a block are taken before moving on to the next block. The order in which the blocks are run should also be randomized if possible. This procedure achieves the objective of minimizing within-block variability while any time-related effects are isolated indirectly through the blocking variable.

The classical statistical analysis technique for a randomized complete block design is the two-way crossed classification ANOVA, introduced in Section 5.4, where “blocks” are the second factor. The model is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \quad (7.2)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, n$$

where

y_{ijk} = the observation of the k^{th} replicate at the i^{th} level of the factor of interest in the j^{th} block

μ = the overall mean

α_i = the effect of the i^{th} level of the factor of interest

β_j = the effect of the j^{th} block

$\alpha\beta_{ij}$ = the interaction of the i^{th} level of the factor and the j^{th} block

ϵ_{ijk} = the random experimental error affecting y_{ijk}

In this formulation, there are $n \geq 1$ replicate experimental units (or $n \geq 1$ replicate measurements on a single experimental unit) at each factor level within each block. That is, a balanced, complete experiment (Section 5.4.1) is assumed. Methods for analyzing data from unbalanced and/or incomplete experiments are not presented in this text.

Note that there are several possible ways of analyzing the data, depending upon whether the factor of interest and the blocking factor are fixed or random effects factors. If the experiment is not replicated (if $n = 1$), then an assumption of no interaction is required to test the effect of the factor of interest when both factors are fixed or when the factor of interest is random and the blocking factor is fixed. The assumption of no interaction is not required to test the effect of the factor of interest if both factors are random or if the factor of interest is fixed and the blocking factor is random. If the experiment is replicated, no speculative assumptions about the interaction need to be made because the interaction effect can be tested. A test for a blocking factor effect is usually of little interest because the blocks are structured and used in a nonrandom fashion. See Schefflé (1959) for a further discussion of this topic. Snedecor and Cochran (1980) present a technique for determining whether blocking is effective compared to the completely randomized design.

The paired t-test (Section 4.4) is a special case of the randomized complete block ANOVA, where there are only two levels of the factor of interest and each pair of observations defines a block. The interested reader can refer to Montgomery (1976), Cochran and Cox (1957), or Hicks (1982) for additional discussions of randomized complete block designs. Nonparametric analysis techniques for the randomized complete block design are discussed in Section 9.4.1. An example of a randomized complete block design is now considered.

Example 7.2 For the nondestructive instrument problem described in Example 7.1, assume that only seven pellets are to be measured, with each pellet being measured by each of the three instruments. Set up a random-

ized complete block design to compare the three instruments. For instructional purposes, the data given in Table 7.2 are utilized, where each row now corresponds to the measurements taken on a single pellet.

The pellets are the blocks in this experiment. If each of the seven pellets is measured once by all three instruments, the design has complete blocks. Assuming that the pellets (blocks) are a random sample of pellets from a batch, they are measured in the order in which they were selected from the batch. This is equivalent to randomizing the run order of the blocks. Then the order in which the three instruments are used to measure a single pellet is randomized. Using the y_{ij} notation, one possible randomized run order is displayed in Table 7.4.

TABLE 7.4
One Possible Randomized Run Order for a
Randomized Complete Block Design

Block (pellet) run order*	Order instruments are used	Measured response
1	3	y_{31}
	1	y_{11}
	2	y_{21}
2	1	y_{12}
	3	y_{32}
	2	y_{22}
3	3	y_{33}
	2	y_{23}
	1	y_{13}
4	2	y_{24}
	1	y_{14}
	3	y_{34}
5	1	y_{15}
	2	y_{25}
	3	y_{35}
6	1	y_{16}
	3	y_{36}
	2	y_{26}
7	3	y_{37}
	2	y_{27}
	1	y_{17}

*Assume that pellets are randomly selected from a batch of pellets.

The appropriate model is Equation 7.2, where α_i is the effect of the i^{th} instrument, β_j is the effect of the j^{th} pellet, $\alpha\beta_{ij}$ is the interaction effect of the i^{th} instrument and j^{th} pellet, and ϵ_{ijk} is the effect of random experimental error on the k^{th} measurement taken on the j^{th} pellet with the i^{th} instrument ($n = 1$ in this example).

The appropriate method of statistical analysis is the two-way crossed classifications analysis of variance, presented in Section 5.4, where instruments is a fixed effects factor and the blocking factor, pellets, is a random effects factor. The ANOVA results for this example are displayed in Table 7.5. The entries are calculated as discussed in Section 5.4.1. Note that because this experiment was not replicated, the experimental error and interaction effects cannot be separately estimated.

TABLE 7.5
Two-Way ANOVA Table for a
Randomized Complete Block Design

Source	df	SS	MS	F
Instruments	2	0.00050	0.00025	0.42
Blocks (pellets)	6	0.01253	0.00209	
Error	12	0.00704	0.00059	
Total	20	0.02007		

As seen from the SS and MS columns in Table 7.5, a considerable portion of the total variability is due to the variability between blocks (pellets). Thus, the error mean square is reduced, and a larger F-ratio is obtained for testing the hypothesis of instrument equivalence than in Example 7.1, where the between-pellet variability is not isolated; however, because $F = 0.42 < F_{0.95}(2,12) = 3.89$, the observed differences among the three instruments are not statistically significant.

7.2.3 Latin Square Designs

The previous section illustrates that the randomized complete block design isolates one extraneous source of variability in the experimental units by imposing a single restriction on the randomization; specifically, each level of the factor of interest occurs once within each block. **Latin square designs** isolate two extraneous sources of variability through the use of two blocking factors. Interest is still centered on one factor, but the two

blocking factors and their levels are chosen so that they account for major sources of variation in the response variable. This results in a more powerful test of the hypothesis that the factor of interest has no effect on the response variable. The Latin square design is not always appropriate because it assumes that there are no two-factor or three-factor interactions among the blocking factors and the factor of interest.

A Latin square design requires that the number of levels of both blocking factors be the same as the number of levels of the factor of interest. If this common number of levels is represented by p , the resulting design is called $p \times p$ Latin square design. For a Latin square, capital letters A, B, C, ... are often used to represent the levels of the factor of interest.† For example, one possible 3×3 Latin square is represented by

Levels of 2nd blocking factor	Levels of 1st blocking factor		
	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

The blocks might be determined by such factors as batch of material, method of analysis, laboratories, or day of the week.

Note in the 3×3 example above, that the distinctive feature of a Latin square design is that each letter appears exactly once in each row and in each column. This feature defines the two restrictions on randomization that isolate two extraneous sources of variability. It may appear that there is no place for randomization of the experimental units in a Latin square design. However, randomization does have a place and is implemented by randomly selecting one of the many possible squares of a given size. As seen in Table 7.6, the number of possible Latin squares of a particular size increases quite rapidly as the size increases; thus, the random selection of a Latin square may not be a simple task. Fisher and Yates (1953) discuss a procedure for directly generating a random Latin square design. Cochran and Cox (1957) also discuss how to select a random Latin square design.

The principle of randomization is also applied in the specification of the run order for the individual trials. For a Latin square design, the

†This convention is common in texts on experimental design. It differs, however, from the usual notation for multifactor experiments, where capital letters are often used to represent the factors while integers are used to represent the levels of each factor.

TABLE 7.6
Standard Squares and Numbers of Latin
Squares of Various Sizes†‡

Size	3 × 3	4 × 4	5 × 5	6 × 6	p × p
Example of a standard square	ABC BCA CAB	ABCD BCDA CDAB DABC	ABCDE BAECD CDAEB DEBAC ECDBA	ABCDEF BCFADE CFBEAD DEABFC EADFCB FDECBA	ABC ... P BCD ... A CDE ... B . . . PAB ... (P - 1)
Number of standard squares	1	4	56	9408	-
Total number of Latin squares	12	576	161,280	8,128,851,200	$p!(p-1)! \times$ (number of standard squares)

†This table is adapted from one presented by Montgomery (1976).

‡A standard Latin square is one in which the first row and column contain the letters in alphabetical order.

$p \times p = p^2$ trials should be run in completely random order if possible. This ensures that any uncontrolled sources of variability are evenly distributed over the experiment.

The appropriate statistical analysis procedure for a Latin square design is a three-way analysis of variance, where all interactions are assumed to be negligible. This is an extension of the two-way ANOVA presented in Section 5.4.1. Computational details of the three-way ANOVA are not presented in this text, however. The model for this design is given by

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkl} \quad (7.3)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, a$$

$$k = 1, 2, \dots, a$$

$$l = 1, 2, \dots, n$$

where

y_{ijkl} = l^{th} replicate observation at i^{th} level of the factor of interest and at j^{th} and k^{th} levels of the two blocking factors

μ = overall mean

α_i = effect of i^{th} level of the factor of interest

β_j = effect of j^{th} level of the first blocking factor

γ_k = effect of k^{th} level of the second blocking factor

ϵ_{ijkl} = the random experimental error affecting y_{ijkl}

Although the model indicates that replication is possible with a Latin square design, it is not required for hypothesis testing because all interactions are assumed to be negligible. The development and statistical analysis of a Latin square design is illustrated in the following example.

Example 7.3 In a particular fuel fabrication facility, three different NDA instruments are used to measure the ^{235}U content of cans of highly enriched uranium scrap. At any one time, any one of three operators may use any one of the NDA instruments. Set up a Latin square design to determine whether any significant differences exist among the three instruments.

“Instruments” is the factor of interest, while “operators” and “cans of scrap” are the two blocking factors. Three cans of scrap are measured by each operator with each instrument to yield a 3×3 Latin square design. The three cans of scrap are the experimental units; they are randomly assigned to the three rows of the Latin square. The three operators are randomly assigned to the three columns of the Latin square. The three instruments are randomly assigned to the designations A, B, and C. Then one of the 12 possible 3×3 Latin squares is randomly selected. Suppose the following square is selected:

Can of scrap	Operator		
	1	2	3
1	B	A	C
2	C	B	A
3	A	C	B

For example, the “B” in the upper left hand corner of the square indicates that the first operator uses the second NDA instrument to measure the first can of scrap.

The order in which each of the nine trials is performed is completely randomized. One such possibility is given by the following table:

Can of scrap	Operator		
	1	2	3
1	8	5	2
2	6	3	7
3	4	1	9

Suppose that the following results (in kilograms of ^{235}U) are obtained:

Can of scrap	Operator		
	1	2	3
1	12.42	12.31	12.68
2	13.75	13.62	13.50
3	11.70	12.13	11.91

The results from a three-way analysis of variance are displayed in Table 7.7. A major portion of the total variability in the data is due to differences among cans of scrap. The different operators also account for

TABLE 7.7
Three-Way ANOVA Table for Latin Square Design

Source	df	SS	MS	F
Instruments	2	0.185356	0.092678	170.36
Cans of scrap	2	4.564156	2.282078	
Operators	2	0.009489	0.004745	
Error	2	0.001088	0.000544	
Total	8	4.760089		

some of the total variation. The F-ratio for differences among instruments is $F = MS_{\text{Instruments}}/MS_{\text{Error}} = 170.36$ and is significant at the $\alpha = 0.01$ level because it exceeds $F_{0.99}(2,2) = 99.00$ obtained from Table A6.

Note that the Latin square design is also appropriate for problems where three noninteracting factors with the same number of levels are of interest; that is, it is not necessary that two of the factors be blocking factors. The three-way ANOVA procedure is computationally the same; it partitions the total sum of squares into components due to the three factors, regardless of whether they are independent variables being studied or blocking factors. Note also that the Latin square design is a special case of the fractional factorial designs discussed in Section 7.4.

7.2.4 Other Designs

This section briefly introduces several other experimental designs for a single factor experiment, with references for the interested reader.

A **Graeco-Latin square design** isolates three extraneous sources of variability by using three blocking factors. The factor of interest and the three blocking factors must all have the same number of levels. This design is discussed by Montgomery (1976), Cochran and Cox (1957), and Hicks (1982), among others.

A **randomized incomplete block design** is one where there are more treatments (levels of the factor of interest) than can be run in a single block. Discussion of these designs is provided by Montgomery (1976), Hicks (1982), Cochran and Cox (1957), and Fisher and Yates (1953).

Youden square designs are incomplete Latin square designs in which the number of levels of one blocking factor is less than the number of levels of the other blocking factor and the factor of interest. A table of Youden square designs is provided by Davies (1956). Discussion of Youden square designs is provided by Montgomery (1976), Hicks (1982), and Cochran and Cox (1957).

7.3 FACTORIAL DESIGNS FOR MULTIFACTOR EXPERIMENTS

In Section 7.2, all of the experimental designs considered were for investigating the effect of a single factor on a measured response variable. In this section, **factorial designs** are introduced for the simultaneous study of the effects of two or more factors. In a factorial design, if there are a levels of factor A, b levels of factor B, and c levels of factor C, the experiment will include all of the (a) (b) (c) different combinations of the factor levels. Because of this multiplicative method of counting the number of possible factor level combinations, the notation $a \times b \times c$ is often used to represent a particular factorial design.

Factorial experiments are more efficient in the use of experimental resources than experiments that investigate several factors one at a time. For the same amount of effort, factorial designs provide information about factor interactions that cannot be obtained by investigating factors one at a time. Thus, the designs discussed in this section are preferred over one-at-a-time investigation of each factor using designs such as those discussed in Section 7.2.

7.3.1 Effects and Interactions

The effect† of a factor is defined as the change in response produced by a change in the level of the factor. Because more than one factor is

†The term "effect" is often used interchangeably for both fixed and random effects. This convention is followed here; however, for instructional purposes, the present discussion applies only to fixed effects.

being varied in a factorial experiment, there are three basic categories of effects that can be evaluated: they are **main effects**, **simple effects**, and **interaction effects**. These are defined and illustrated in the following discussion.

Consider the results of the 2×2 factorial experiment displayed in Table 7.8.

TABLE 7.8
Results of a 2×2 Factorial Experiment

		Factor A	
		1	2
Factor B	1	15	25
	2	17	26

The main effect of factor A is the difference between the average response at the first level of A and the average response at the second level of A, which is computed as

$$\text{Main effect of A} = \frac{(25 + 26)}{2} - \frac{(15 + 17)}{2} = 9.5$$

That is, changing factor A from level 1 to level 2 has resulted in an average increase of 9.5 in the measured response. Similarly, the main effect of factor B is computed as

$$\text{Main effect of B} = \frac{(17 + 26)}{2} - \frac{(15 + 25)}{2} = 1.5$$

That is, changing factor B from level 1 to level 2 has resulted in an average increase of 1.5 in the measured response.

In some experiments, the difference in response between the levels of one factor is not the same at all levels of the other factors. This occurs when there are **interactions** among the factors; that is, the factors interact in their effect on the response variable. Consider the results of the 2×2 factorial experiment displayed in Table 7.9. At the first level of factor B, the simple effect of factor A is

$$\text{Simple effect of A at } B_1 = 20 - 10 = 10$$

TABLE 7.9

Results of a 2×2 Factorial Experiment with Interaction and Nonsignificant Main Effects

		Factor A	
		1	2
Factor B	1	10	20
	2	25	8

At the second level of B, the simple effect of factor A is

$$\text{Simple effect of A at } B_2 = 8 - 25 = -17$$

Because the effect of factor A depends on the level of factor B, there is an interaction between factors A and B. Note that the simple effects of factor B are

$$\text{Simple effect of B at } A_1 = 25 - 10 = 15$$

and

$$\text{Simple effect of B at } A_2 = 8 - 20 = -12$$

Clearly, the effect of factor B depends on the level of A.

When two or more factors interact, their main effects may be meaningless, and extreme caution must be used in interpreting analysis of variance results. For example, using the data from Table 7.9, the main effects are

$$\text{Main effect of A} = \frac{(20 + 8)}{2} - \frac{(10 + 25)}{2} = -3.5$$

and

$$\text{Main effect of B} = \frac{(25 + 8)}{2} - \frac{(10 + 20)}{2} = 1.5$$

Depending on the size of the experimental error variance, these main effects might be deemed nonsignificant. Then it might be tempting to conclude that factors A and B have no effect on the measured response; however, examination of the simple effects indicates that such a conclusion is erroneous. Thus, in this example, knowledge of the interaction and exami-

nation of the simple effects is more useful than knowledge of only the main effects.

Although interactions may tend to mask main effects, they do not always preclude interpretation of main effects. Consider the data in Table 7.10. The simple effects are

$$\text{Simple effect of A at } B_1 = 12 - 5 = 7$$

$$\text{Simple effect of A at } B_2 = 20 - 30 = -10$$

$$\text{Simple effect of B at } A_1 = 30 - 5 = 25$$

and

$$\text{Simple effect of B at } A_2 = 20 - 12 = 8$$

and the main effects are

$$\text{Main effect of A} = \frac{(12 + 20)}{2} - \frac{(5 + 30)}{2} = -1.5$$

and

$$\text{Main effect of B} = \frac{(30 + 20)}{2} - \frac{(5 + 12)}{2} = 16.5$$

The simple effects indicate that there is an interaction. The main effect of factor A, -1.5 , is quite small relative to the simple effects and may be negligible, depending on the size of the experimental error variance. The main effect of factor B, 16.5 , is meaningful, however. This illustrates that the presence of interaction does not always preclude interpretation of main effects.

TABLE 7.10

Results of a 2×2 Factorial Experiment with Interaction and a Significant Main Effect

		Factor A	
		1	2
Factor B	1	5	12
	2	30	20

Effects and interactions are sometimes more easily visualized by considering plots of the data, such as those in Figs. 7.1, 7.2, and 7.3 for the data from Tables 7.8, 7.9, and 7.10, respectively. The slope of each line in the plots represents a simple effect, distances between lines (at their mid-

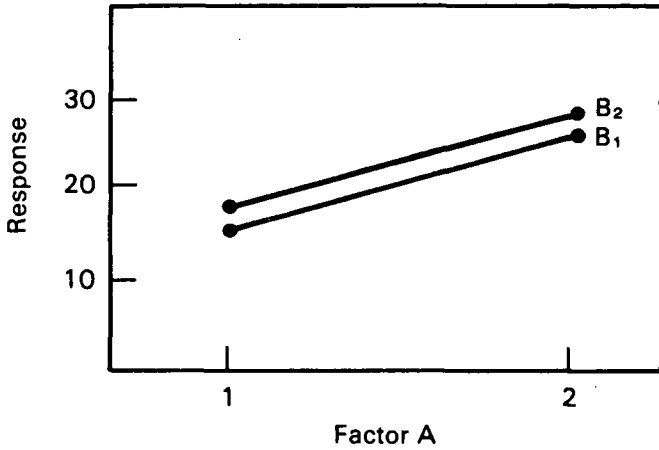


Figure 7.1 A 2×2 factorial experiment without interaction.

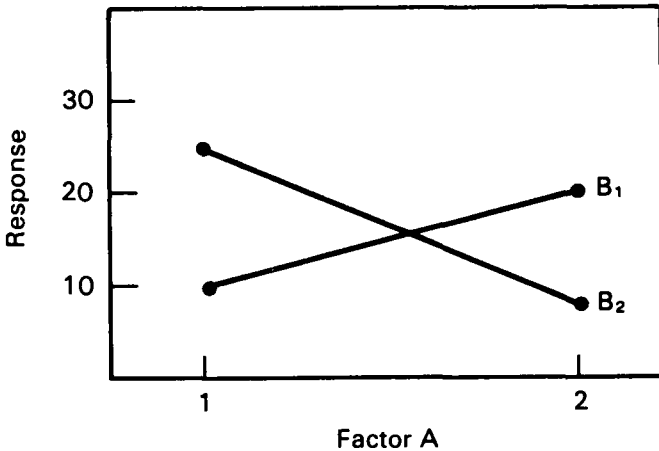


Figure 7.2 A 2×2 factorial experiment with interaction—main effects not meaningful.

points) represent main effects, and the parallelism (or lack thereof) of lines within a plot indicates the degree of interaction between the two factors.

The calculation of effects and the data plots illustrated in the previous example can readily be extended to more complex factorial designs. Note, however, that these techniques are not intended to provide objective tests of significance for main effects or interaction effects. It is necessary to use analysis of variance techniques, such as those discussed in Chapter 5, to analyze the data from factorial designs and to perform hypothesis tests on

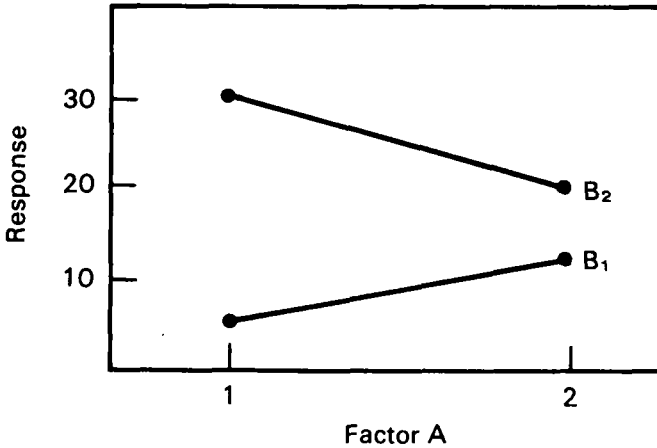


Figure 7.3 A 2×2 factorial experiment with interaction—main effect of factor B meaningful.

main effects and interaction effects. Note that a factorial design must be replicated to allow complete testing of all interactions. If higher-order interactions can be assumed negligible, however, their sums of squares and degrees of freedom can be combined to provide an estimate of the experimental error variance that can be used to test the significance of main effects and lower-order interactions.

7.3.2 2^k and 3^k Designs

In many types of multifactor experiments, interest centers on investigating only two or three levels of each factor. The multiplicative notation $2 \times 2 \times \dots \times 2$ and $3 \times 3 \times \dots \times 3$ naturally leads to the shorter exponential notation 2^k and 3^k , where k is the number of factors. The 2^k and 3^k designs are valid for both fixed and random effects factors. They are k -way crossed classifications, which are extensions of the two-way crossed classifications discussed in Section 5.4. For factors with fixed effects in a 2^k design, the levels are often chosen to bound the range of interest for each factor and are referred to as “high” and “low” levels. A 2^k fixed effects design is adequate when the effects of the factors are linear; that is, as the level of a factor is increased, the response variable either increases or decreases as a linear function of the factor levels. A 2^k design also provides information for isolating and testing two-factor, three-factor, and up through k -factor interactions. For factors with fixed effects in a 3^k design, an intermediate value of each factor (in addition to high and low values) is included. A 3^k fixed effects design allows the investigation of curvature in the relationship between the response variable and

the factor levels. For random factors in these designs, the levels are randomly selected from a population of possible levels, and the purpose of the experiment is to estimate and test for significance of the variance components.

The statistical analysis procedure for a 2^k or 3^k factorial experiment is the k -way analysis of variance, which is an extension of the two-way analysis of variance discussed in Section 5.4. Regression analysis of fixed effects factorial designs is also appropriate. The emphasis of the following two subsections is on the designs themselves and not the method of analysis.

7.3.2.1 2^k Designs

The 2^k factorial design is appropriate for investigating the effects of k factors on a response variable, where only two levels of each factor exist or are of interest. It is also appropriate when the factors under study have several (or a continuum) of possible levels, but only a broad overview of the situation (such as that provided by investigating only two levels of each factor) is required. One example of this second area of application is a screening experiment, where the purpose is to determine which factors and interactions are important for further investigation in a more detailed study. Another example is a regression application, where it is known that the relationship between the response variable and each of the factors is linear. (If it is not linear or if a test for nonlinearity is desired, a 3^k design or other design with more than two levels of each factor is often used.) In this case, interactions among the factors indicate that the linear relationship between the response variable and the levels of a given factor changes with the levels of the other factors. This is illustrated above in Figs. 7.2 and 7.3.

Table 7.11 displays 2^k factorial design layouts for $k = 2, 3,$ and 4 . In performing these experiments, random assignment of experimental units to the factor level combinations is required, and the order in which the trials are run must be randomized. Replication of at least one factor level combination (and preferably the entire experiment) is required to investigate all main effects and interactions. With no replication, it is possible to statistically test the hypotheses of no main effects or lower-order interactions if some of the higher-order interactions are assumed to be negligible. This practice is not generally recommended, however, unless prior experimentation or scientific theory supports the nonexistence of the higher-order interactions.

Table 7.12 lists the number of trials required for one complete replicate of each 2^k design for $k = 2, 3, \dots, 10$. For five or more factors, the number of trials in a 2^k design is rather large. Designs based on two levels per factor, but which involve fewer trials, are discussed in Section 7.4.1.

TABLE 7.11
Layouts for 2^2 , 2^3 , and 2^4 Factorial Designs†

Trial	Factors			
	A	B	C	D
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4 = 2^2	+	+	-	-
<hr/>				
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8 = 2^3	+	+	+	-
<hr/>				
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16 = 2^4	+	+	+	+

†The first (or low) level of a factor is represented by “-” and the second (or high) level by “+”. This abbreviated notation is commonly used in experimental design texts and literature.

TABLE 7.12
Number of Trials for One Replicate of a 2^k Design

Number of factors (k)	Number of trials (2^k)
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024

Example 7.4 A laboratory manager is asked to investigate a certain laboratory procedure that is always performed by one of two technicians using one of two instruments. Design a 2^2 factorial experiment to estimate and test for significance of the main effects due to instruments and technicians and the effect of interaction between instruments and technicians.

Because only the two technicians and two instruments are of specific interest, the effects of technician and instrument are fixed effects. Although a single replicate of a 2^2 design is sufficient to estimate the main effects and two-factor interaction for instruments and technicians, replication is required to perform statistical tests. Suppose that a 2^2 design replicated three times is chosen. Referring to Table 5.10 in Section 5.4.2, this will yield an estimate of the experimental error variance σ^2 with 8 degrees of freedom. Assume that a single sample of experimental material is divided into 12 aliquots which are randomly assigned to the 12 trials listed in Table 7.13. The design (including randomized run order) is summa-

TABLE 7.13

 2^2 Design Replicated Three Times

Trial	Instrument	Technician	Random run order
1	1	1	10
2	2	1	7
3	1	2	11
4	2	2	5
5	1	1	8
6	2	1	4
7	1	2	3
8	2	2	9
9	1	1	1
10	2	1	6
11	1	2	12
12	2	2	2

riized in Table 7.13. The two-way analysis of variance methods for the fixed effects model, presented in Section 5.4.2, would be used to analyze the data from this experiment.

The “-” and “+” notation introduced in Table 7.11 can be utilized in a convenient procedure for estimating and testing main effects and interaction effects from a balanced 2^k factorial experiment, where all factors have fixed effects. To introduce and illustrate the procedure, consider some hypothetical data from a 2^3 factorial experiment with two replicates at each combination of the factor levels. The factor levels, the corresponding replicate values, and the totals and averages of the replicates are arranged in the format shown in Table 7.14.

TABLE 7.14

Example of Data from a 2^3 Experiment

A	B	C	Replicates	Totals	Averages
-	-	-	13, 10	23	11.5
+	-	-	19, 23	42	21.0
-	+	-	2, 6	8	4.0
+	+	-	10, 7	17	8.5
-	-	+	14, 17	31	15.5
+	-	+	11, 7	18	9.0
-	+	+	4, 1	5	2.5
+	+	+	13, 9	22	11.0

The “-” and “+” are then treated as -1 and 1, and Table 7.14 is expanded by constructing interaction columns as shown in Table 7.15. For example, the elements of the AB column are computed by multiplying the corresponding elements of the A and B columns.

TABLE 7.15

Expanded 2^3 Data Table

A	B	C	AB	AC	BC	ABC	Averages
-	-	-	+	+	+	-	11.5
+	-	-	-	-	+	+	21.0
-	+	-	-	+	-	+	4.0
+	+	-	+	-	-	-	8.5
-	-	+	+	-	-	+	15.5
+	-	+	-	+	-	-	9.0
-	+	+	-	-	+	-	2.5
+	+	+	+	+	+	+	11.0

The main effects and interaction effects are estimated by assigning the “-” and “+” signs in a given column to the corresponding averages and

then summing. The sum is then divided by the number of “+” signs. For example, the estimated main effect of A is computed as

$$\begin{aligned}\text{Main effect of A} &= \frac{-11.5 + 21 - 4 + 8.5 - 15.5 + 9 - 2.5 + 11}{4} \\ &= \frac{16}{4} = 4.0\end{aligned}$$

The estimated ABC interaction effect is computed as

$$\begin{aligned}\text{ABC interaction effect} &= \frac{-11.5 + 21 + 4 - 8.5 + 15.5 - 9 - 2.5 + 11}{4} \\ &= \frac{20}{4} = 5.0\end{aligned}$$

The estimated effects are displayed in Table 7.16.

TABLE 7.16

**Estimated Effects for
2³ Example**

Source	Estimated effect
A	4.00
B	-7.75
C	-1.75
AB	2.50
AC	-3.00
BC	2.25
ABC	5.00

To statistically test the hypotheses that the true main effects and interaction effects are zero, an estimate of the experimental error variance σ^2 is required. One of the underlying assumptions in the analysis of variance is that the experimental error variance is constant over all factor level combinations. When this assumption is valid, and when the experiment has been replicated at least twice, an estimate of the experimental error variance is computed by applying the one-way analysis of variance formulas presented in Section 5.1.1. For the present example, the eight factor level

combinations are treated as the classes in the one-way analysis of variance context. There are $2^k = 8$ classes with $n = 2$ replicate values in each class. Let y_{ij} denote the j^{th} replicate in the i^{th} class. Then, Equation 5.20 is applied to the replicate data and the totals displayed in Table 7.14, which gives

$$\begin{aligned} SS_{\text{within}} &= \sum_{i=1}^8 \sum_{j=1}^2 y_{ij}^2 - \frac{1}{2} \sum_{i=1}^8 y_i^2 \\ &= (13^2 + 10^2 + \dots + 13^2 + 9^2) - \frac{1}{2} (23^2 + 42^2 + \dots + 22^2) \\ &= 2270 - \frac{1}{2} (4440) = 50 \end{aligned}$$

The degrees of freedom are

$$df = 2^k(n - 1) = 8(2 - 1) = 8$$

Thus, the estimate of the experimental error variance is

$$\hat{\sigma}^2 = MS_{\text{within}} = 50/8 = 6.25$$

Next, note that each average in Table 7.14 has variance $\sigma^2/n = \sigma^2/2$. Because each effect estimate in Table 7.15 is a linear combination of the $2^k = 8$ averages, with all of the linear coefficients being either $-1/4$ or $1/4$, the variance of each estimate is

$$\sum_{i=1}^8 \left(\frac{1}{4}\right)^2 (\sigma^2/2) = 8 \left(\frac{1}{4}\right)^2 (\sigma^2/2) = \frac{\sigma^2}{4}$$

This is estimated by

$$\frac{\hat{\sigma}^2}{4} = \frac{6.25}{4} = 1.5625$$

and the estimated standard deviation of each effect estimate is

$$\left(\frac{\hat{\sigma}^2}{4}\right)^{1/2} = \sqrt{1.5625} = 1.25$$

Note that for any complete 2^k factorial experiment with n replicate values at each combination of the factor levels, the variance of each effect estimate is

$$\text{Var}(\text{effect estimate}) = \frac{\sigma^2}{n2^{k-2}} \quad (7.4)$$

A test statistic t is computed for each of the effect estimates in Table 7.15 by letting

$$\begin{aligned} t &= \frac{\text{Effect estimate}}{\text{Estimated standard deviation of effect estimate}} \\ &= \frac{\text{Effect estimate}}{1.25} \end{aligned} \quad (7.5)$$

For a specified significance level α , the hypothesis of no true effect is rejected if either

$$t \leq -t_{1-\alpha/2}[2^k(n-1)]$$

or (7.6)

$$t \geq t_{1-\alpha/2}[2^k(n-1)]$$

where $t_{1-\alpha/2}[2^k(n-1)]$ is obtained from Table A5.

For the present example, let $\alpha = 0.05$. Then, $t_{0.975}(8) = 2.31$ is obtained from Table A5. A summary of the computed test statistics is given in Table 7.17. Note that factors A and B have estimated main

TABLE 7.17
Hypothesis Test Results

Source	Estimated effect	Test statistic	Reject H_0
A	4.00	3.20	Yes
B	-7.75	-6.20	Yes
C	-1.75	-1.40	No
AB	2.50	2.00	No
AC	-3.00	-2.40	Yes
BC	2.25	1.80	No
ABC	5.00	4.00	Yes

effects that differ significantly from zero at the 0.05 level of significance. Also, the estimated AC and ABC interaction effects are significantly different from zero at the 0.05 level. The conclusion is that the corresponding

true effects are probably not zero. The interpretation of this analysis should be supported by interaction plots such as those shown in Figs. 7.1, 7.2, and 7.3.

7.3.2.2 3^k Designs

The 3^k factorial design is appropriate for investigating the effects and interactions of k factors, where three levels of each factor are considered. It may be that only three levels exist or that three (out of several or a continuum) are chosen to be studied because previous experimentation or physical laws suggest that the relationship between the response variable and the factors is nonlinear in the factor levels. With fixed effects, including three levels of each factor provides adequate data to estimate and test for significance of both linear and quadratic (second-order curvature) relationships between the response variable and the factor levels. The interpretation of interactions is more complicated than for the 2^k designs, and data plots—analogueous to Figures 7.1, 7.2, and 7.3—are valuable aids. The low and high levels of each factor are analogueous to those discussed for the 2^k designs in the previous section. It is desirable to choose the intermediate level of each quantitative factor to be halfway between the high and low levels. For some applications, the halfway point is chosen on a logarithmic scale.

Table 7.18 displays the layouts for the 3^2 and 3^3 factorial designs. The number of trials for one replicate (3^k) increases substantially as the number of factors increases, as shown in Table 7.19. Designs that are based on three levels per factor but involve fewer trials than the 3^k designs are called fractional designs and are briefly discussed in Section 7.4.2.

In performing any 3^k experiment, randomization must be employed both in assigning experimental units to the trials and in determining the run order of the trials. Replication of at least one of the factor level combinations (and preferably the entire experiment) is required to statistically test the significance of all main effects and interaction effects. When there is no replication, main effects and lower-order interaction effects can be statistically tested if some higher-order interaction effects can be assumed to be negligible. This practice is not generally recommended, however, unless prior experimentation or scientific theory supports the nonexistence of the higher-order interaction effects.

Example 7.5 In the recovery of uranium from spent reactor fuel, a solvent extraction process is used to remove uranium from the dissolved fuel, leaving the fission products behind. Factors believed to be of importance in this process are percent by volume of extractant and acidity of the solution. The levels of the factors to be investigated are as follows:

Factor	Level	
% Extractant	Low	10
	Intermediate	20
	High	30
Acid molarity	Low	2
	Intermediate	5
	High	8

TABLE 7.18
Layouts for 3^2 , 3^3 , and Factorial Designs†

Trial	Factors		
	A	B	C
1	—	—	—
2	0	—	—
3	+	—	—
4	—	0	—
5	0	0	—
6	+	0	—
7	—	+	—
8	0	+	—
9 = 3^2	+	+	—
10	—	—	0
11	0	—	0
12	+	—	0
13	—	0	0
14	0	0	0
15	+	0	0
16	—	+	0
17	0	+	0
18	+	+	0
19	—	—	+
20	0	—	+
21	+	—	+
22	—	0	+
23	0	0	+
24	+	0	+
25	—	+	+
26	0	+	+
27 = 3^3	+	+	+

†The first (or low) level of a factor is represented by “—”, the second (or middle) level by “0”, and the third (or high) level by “+”.

TABLE 7.19
Number of Trials for One
Replicate of a 3^k Design

Number of factors (k)	Number of trials (3^k)
2	9
3	27
4	81
5	243
6	729
7	2187
8	6561

Develop a 3^2 factorial design for testing the effects of the two factors and their interaction.

A 3^2 design with two replicates is displayed in Table 7.20. Trials 1 through 9 are the first replicate, and trials 10 through 18 are the second replicate. Referring to Table 5.10, the replication will yield an estimate of the experimental error variance σ^2 with 9 degrees of freedom. The two-

TABLE 7.20
 3^2 Design Replicated Twice

Trial	% extractant	Acid molarity	Random run order
1	10	2	3
2	20	2	17
3	30	2	5
4	10	5	10
5	20	5	1
6	30	5	14
7	10	8	16
8	20	8	8
9	30	8	13
10	10	2	9
11	20	2	11
12	30	2	7
13	10	5	12
14	20	5	6
15	30	5	4
16	10	8	18
17	20	8	2
18	30	8	15

way analysis of variance methods for the fixed effects model, presented in Section 5.4.2, would be used to analyze the data from this experiment.

7.3.3 Mixed Level Factorial Designs

For some experimental situations, a design with an equal number of levels for all factors, such as a 2^k or 3^k design, is not appropriate. This poses no problem because factorial designs are quite general, only requiring at least two factors and at least two levels per factor. Recall that a factorial experiment is one in which all levels of each factor are combined with all levels of all other factors. That is, if there are k factors and the i^{th} factor has a_i levels, then a single replicate of the $a_1 \times a_2 \times \dots \times a_k$ factorial experiment consists of $(a_1)(a_2)(a_3) \dots (a_k)$ factor level combinations. For example, in an experiment with $k = 3$ factors, suppose that two factors have 3 levels and the other factor has 4 levels. Then a single replicate of the $3 \times 3 \times 4$ experiment has $(3)(3)(4) = 36$ factor level combinations. The principle of randomization should be applied in assigning experimental units to the trials and in determining the run order of the trials during the experiment. A mixed-level factorial design is illustrated in the following example.

Example 7.6 Referring to Example 7.5, suppose that a fourth level of the percent extractant factor (40%) is to be included in the experiment. The layout for a single replicate of a 4×3 design is displayed in Table 7.21.

A considerable amount of literature exists on the subject of factorial designs, too much to be exhaustively referenced here. Interested readers can refer to such texts as Box et al. (1978), Montgomery (1976), Hicks (1982), Cochran and Cox (1957), or Davies (1956) for additional general discussions of the design and analysis of factorial experiments.

7.4 FRACTIONAL FACTORIAL DESIGNS

As the number of factors or the number of levels per factor increases in a factorial design, the number of trials required for a complete replicate of the design can rapidly outgrow the available resources of an experimenter. Tables 7.12 and 7.19 illustrate this point for the 2^k and 3^k factorial designs.

TABLE 7.21
Single Replicate of a 4×3 Design

Trial	% extractant	Acid molarity
1	10	2
2	20	2
3	30	2
4	40	2
5	10	5
6	20	5
7	30	5
8	40	5
9	10	8
10	20	8
11	30	8
12	40	8

In some experimental situations, it is reasonable to assume that some of the interaction effects are negligible. Often, in these situations, main effects and nonnegligible interaction effects can be estimated and tested by performing only a fraction of the complete factorial experiment. **Fractional factorial designs** are the topic of this section.†

The reduction in the size of an experiment afforded by a fractional factorial design has its consequences. Because fewer trials of the experiment are being run, the precision of main effect, interaction effect, or variance component estimates is reduced. Also, to estimate and test some effects, it is necessary to assume that other specific effects are negligible. Estimates of the effects of interest may be misleading if any of the effects assumed to be negligible are in fact significant. This is because fractional factorial designs are chosen in such a way that two or more effects will have the same estimator (within multiplication by -1). Such effects are indistinguishable from each other because they cannot be isolated and separately estimated and tested. Fixed effects that have the same estimator (within multiplication by -1) are said to be **confounded** or **aliased**. Consequently, effects that are confounded are called **aliases**. The number of confounded effects for a certain fractional factorial design depends upon the number of factors and the fraction of the full factorial chosen.

In the following subsections, fractions of 2^k , 3^k , and mixed-level factorial designs are considered. Although a complete discussion of the procedures for generating these fractional designs and determining the associ-

† Although the fixed effects terminology is used in this section, fractional factorial designs are also applicable for estimating variance components for factors with random effects.

ated alias structures is not presented, some illustrative guidelines are given for fractions of 2^k factorial designs. The interested reader is referred to Montgomery (1976), Hicks (1982), Cochran and Cox (1957), or Davies (1956) for the details of the development of these designs and analysis of the resulting experimental data.

7.4.1 Fractions of 2^k Designs

A fractional factorial comprised of a $1/2^p$ fraction of a 2^k design is referred to as a 2^{k-p} fractional factorial design. Table 7.22 displays some summary information for several 2^{k-p} designs. One of the items summarized in Table 7.22 is the best achievable resolution of each size of design.

TABLE 7.22
Summary of Properties of
Several 2^{k-p} Designs

Number of factors (k)	Fraction ($1/2^p$)	Number of trials (2^{k-p})	Best achievable resolution
3	$1/2$	4	III
4	$1/2$	8	IV
5	$1/2$	16	V
	$1/4$	8	III
6	$1/2$	32	VI
	$1/4$	16	IV
	$1/8$	8	III
7	$1/2$	64	VII
	$1/4$	32	IV
	$1/8$	16	IV
	$1/16$	8	III
8	$1/2$	128	VIII
	$1/4$	64	V
	$1/8$	32	IV
	$1/16$	16	IV

This system of classifying designs by their resolution was developed by Box and Hunter (1961a, 1961b) and is explained as follows:

Resolution III Designs. These are designs in which no main effect is aliased with any other main effect, but main effects are aliased with two-factor interactions, and two-factor interactions are aliased with each other.

Resolution IV Designs. These are designs in which no main effect is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with each other.

Resolution \geq V Designs. These are designs in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions may be aliased with three-factor interactions (Resolution V), and so on.

Note, however, that the resolutions displayed in Table 7.22 are the best achievable resolution for a given size fraction of a particular 2^k factorial design. Not all of the possible fractions of a given size have the same resolution. In fact, choice of the defining contrast (discussed below) determines the alias structure and the resolution of the resulting fractional factorial design.

Examples of particular 2^{4-1} , 2^{5-1} , and 2^{5-2} designs and their associated alias structures are displayed in Tables 7.23, 7.24, and 7.25, respectively. In presenting the alias structure in each of these tables, a relationship such as $I = ABCD$ is given. This is technically referred to as a

TABLE 7.23
A Particular 2^{4-1} Design†

Trial	A	B	C	D
1	—	—	—	—
2	+	+	—	—
3	+	—	+	—
4	—	+	+	—
5	+	—	—	+
6	—	+	—	+
7	—	—	+	+
8	+	+	+	+

Aliases ($I = ABCD$)

A = BCD AB = CD
 B = ACD AC = BD
 C = ABD AD = BC
 D = ABC

†One-half replicate of a 2^4 design.

defining contrast and is used to select a fractional factorial design with a particular alias structure. Thus, in Table 7.23 the defining contrast $I = ABCD$ implies that the four-factor interaction $ABCD$ must be assumed to be negligible with respect to experimental error to unbiasedly estimate the overall mean. To unbiasedly estimate and test the main effects, the three-factor interactions must be assumed to be negligible. Three of the six two-factor interactions are estimable if the corresponding aliased two-factor interactions are assumed to be negligible.

TABLE 7.24
 2^{5-1} Design†

Trial	A	B	C	D	E
1	-	-	-	-	-
2	+	+	-	-	-
3	+	-	+	-	-
4	+	-	-	+	-
5	+	-	-	-	+
6	-	+	+	-	-
7	-	+	-	+	-
8	-	+	-	-	+
9	-	-	+	+	-
10	-	-	+	-	+
11	-	-	-	+	+
12	+	+	+	+	-
13	+	+	+	-	+
14	+	+	-	+	+
15	+	-	+	+	+
16	-	+	+	+	+

Aliases ($I = ABCDE$)

A = BCDE	AB = CDE	BD = ACE
B = ACDE	AC = BDE	BE = ACD
C = ABDE	AD = BCE	CD = ABE
D = ABCE	AE = BCD	CE = ABD
E = ABCD	BC = ADE	DE = ABC

†One-half replicate of a 2^5 design.

TABLE 7.25
 2^{5-2} Design†

Trial	A	B	C	D	E
1	-	-	-	-	-
2	+	+	-	-	-
3	-	-	+	+	-
4	+	-	+	-	+
5	-	+	+	-	+
6	+	-	-	+	+
7	-	+	-	+	+
8	+	+	+	+	-

Aliases ($I = ABE = CDE = ABCD$)

A = BE = BCD = ACDE	AC = BD = ADE = BCE
B = AE = ACD = BCDE	AD = BC = ACE = BDE
C = DE = ABD = ABCE	
D = CE = ABC = ABDE	
E = AB = CD = ABCDE	

†One-fourth replicate of a 2^5 design.

TABLE 7.27
One-Half Replicate of a 2^4 Factorial with $I = ABCD$

A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Consider now the one-fourth replicate of a 2^5 factorial given in Table 7.25. The defining contrast is $I = ABE = CDE = ABCD$. Thus, in choosing this design from an expanded table, all rows would be selected that simultaneously have a “+” in the ABE, CDE, and ABCD columns. There are four possible one-fourth replicates, and all of them have the same alias scheme. The four one-fourth replicates can be identified as follows:

One-fourth replicate	ABE	CDE
1	+	+
2	+	-
3	-	+
4	-	-

That is, one of the one-fourth replicates has a “+” in both the ABE and CDE columns, one has a “+” in the ABE column and a “-” in the CDE column, and so on. As with the one-half replicate, there is no advantage in choosing one of the possible one-fourth replicates over the others because they all provide the same information, and they all have the same alias structure.

As with full factorial designs, it is necessary to apply the principle of randomization in assigning experimental units to trials and in determining the run order of the trials of a 2^{k-p} fractional factorial experiment. This is illustrated in Example 7.7.

Example 7.7 Consider a study to investigate the effect on laboratory analysis results due to varying the levels of five factors. The five factors are as follows:

Factor description	Factor name
Sampling method	A
Sample preparation technique	B
Mixing time	C
Instrument	D
Technician	E

The laboratory cannot afford to run $2^5 = 32$ trials, and a design requiring fewer runs is of interest. Past experience indicates that the three-, four-, and five-factor interactions are probably negligible and that the instrument factor (factor D) does not interact with factors A, B, or C. Set up a fractional factorial design for this experiment and outline the ANOVA table.

The 2^{5-1} fractional factorial design shown in Table 7.24 requires only 16 trials. Because it is reasonable to assume that the three-, four-, and five-factor interactions, as well as the AD, BD, and CD interactions, are negligible, all main effects and the remaining two-factor interactions can be estimated and tested as illustrated in the ANOVA table outline in Table 7.28. Note in Table 7.28 that the sums of squares and degrees of

TABLE 7.28
ANOVA Table Outline for 2^{5-1} Design†

Source of Variation	df	SS	MS
A	1		
B	1		
C	1		
D	1		
E	1		
AB	1		
AC	1		
AE	1		
BC	1		
BE	1		
CE	1		
DE	1		
Error	3		
Total	15		

†Example 7.7.

freedom for AD, BD, and CD are combined to provide a sum of squares with three degrees of freedom for error.

In conducting the 2^{5-1} fractional factorial experiment, the trials shown in Table 7.24 are run in random order. Upon collection of the data, the remainder of the ANOVA table in Table 7.28 could be filled in by using the five-way ANOVA, which is an extension of the two-way ANOVA presented in Section 5.4.2.

7.4.2 Fractions of 3^k Designs

The concept of fractional replication also applies to the 3^k designs, where a $1/3^p$ fraction of a 3^k design is denoted by 3^{k-p} . Fractional replication of 3^k designs is especially attractive because a complete replicate can require a large number of trials for even moderate values of k .

The details of how to develop 3^{k-p} designs are not considered here. Texts on experimental design, however, such as Montgomery (1976), Hicks (1982), Cochran and Cox (1957), or Davies (1956), provide the necessary details and guidance for implementing these designs. A publication by Connor and Zelen (1959) includes an extensive selection of 3^{k-p} designs for $4 \leq k \leq 10$.

7.4.3 Fractions of General Factorial Designs

When some or all of the factors in a factorial experiment have more than two levels, the details of determining fractional replicates can be somewhat sophisticated. Because there are so many possible cases, no attempt is made to treat this subject here. Kempthorne (1973) discusses fractions of the general factorial designs, including a brief discussion of mixed level designs. Cochran and Cox (1957) discuss one-half replicates of 3×2^5 and 4×2^4 designs which allow estimation of main effects and all (or nearly all) two-factor interactions.

7.4.4 Blocking with Fractional Replicates

Experimental situations sometimes arise where the experimenter can afford to run a complete factorial design, but it is not possible to run the entire design under homogeneous conditions. For example, suppose that in a given situation a 2^5 factorial design is appropriate. Suppose, however, that it is impossible to run all of the $2^5 = 32$ factor level combinations on the same day and that runs made on different days will use different batches of material. One option available to the experimenter is to run a different fraction of the 32 factor level combinations each day for several days. This situation is not as bleak as it may first appear. If the fractions are properly chosen, it may be possible to salvage most of the information

that would have been obtained from a full factorial experiment run under homogeneous conditions.

The recommended approach to this type of situation is to determine which effects (preferably interaction effects) can be realistically assumed negligible. Then a defining contrast is chosen and is used to split the full factorial into fractions. Each fraction is then run as a block, and the effects that appear in the defining contrast are confounded with the blocks. All other effects are estimable if it can be assumed that they do not interact with the blocks. This is often a reasonable assumption, but it should be carefully considered before implementing this technique.

Suppose that in the situation described earlier, the experimenter can run eight factor level combinations per day, and he is willing to assume that the ABE, CDE, and ABCD interaction effects are negligible. He can then use the defining contrast $I = ABE = CDE = ABCD$ to find the corresponding four one-fourth replicates of the 2^5 factorial design. He would then randomly assign a different one-fourth replicate to each of four days and randomize the order that the factor level combinations are run within a day. Note that, for illustration, the same defining contrast is used here as in Table 7.25.

To determine the four one-fourth replicates, Table 7.29 is constructed. The ABE and CDE columns are examined and the factor level combinations are assigned to blocks as follows: block 1 has all factor level combinations with ABE = "+" and CDE = "+", block 2 has ABE = "+" and CDE = "-", block 3 has ABE = "-" and CDE = "+", and block 4 has ABE = "-" and CDE = "-". Then all factor level combinations in block 1 would be run on one day, all those in block 2 would be run on a different day, and so on. It is important to emphasize that the blocks should be randomly assigned to four different days and the factor level combinations within a block must be run in a randomized order.

In the analysis of variance, the sums of squares and degrees of freedom for ABE, CDE, and ABCD would be pooled to provide a sum of squares for blocks, with three degrees of freedom. The remaining 28 effects, however, could be directly interpreted if it can be assumed that they do not interact with blocks. Because there is no replication, some of the interaction effects would have to be assumed negligible to perform statistical tests. For example, if all four- and five-factor interaction effects are assumed negligible, their sums of squares and degrees of freedom could be pooled to form a sum of squares for experimental error with five degrees of freedom.

This procedure can be extended to accommodate situations where several replicates of a full factorial design must be run, but only a fractional replicate can be run at one time or under homogeneous conditions. Hicks (1982) gives some details and guidelines for the design and analysis of such experiments.

7.5 MULTIFACTOR EXPERIMENTS WITH RANDOMIZATION RESTRICTIONS

Single factor experiments with restrictions on randomization are discussed in Sections 7.2.2, 7.2.3, and 7.2.4. There are also many multifactor experimental situations in which it is either necessary or advisable to restrict randomization.

7.5.1 Randomized Complete Block and Latin Square Designs for Multifactor Experiments

Consider an experiment whose purpose is to simultaneously compare two different sample preparation methods and two different chemical analysis techniques. A 2^2 factorial design (replicated eight times, for example) is used if the experimental material is homogeneous and one technician will perform the whole experiment. Suppose, however, that there is not enough material in one batch to perform all 32 trials of the experiment and that it is necessary to use material from four batches. Here, batch is a blocking factor and restricts complete randomization of the 32 trials. Suppose that two complete replicates of the 2^2 factorial are run using material from a single batch and that these eight trials are completely randomized within each batch. The resulting design is a randomized complete block design with two replicates at each combination of batch, sample preparation method and chemical analysis technique.

Suppose further that not one but four technicians will perform the experiment. Thus, technicians is a second blocking factor. Because the number of levels of the two blocking factors (4) is the same as the number of treatment combinations of the two factors of interest ($2^2 = 4$), a Latin square design for a factorial type experiment could be used. This would require, however, an assumption of no interaction between technicians and batches of material (which seems reasonable) and no interactions between these blocking factors and the factors in the factorial experiment (sample preparation methods and chemical analysis techniques). Two replicates of a 4×4 Latin square would yield the 32 total trials desired.

Details of the development and statistical analysis of multifactor randomized complete block and Latin square designs are found in such texts as Montgomery (1976) or Hicks (1982).

7.5.2 Split-Plot Designs

In some multifactor designs involving randomized complete blocks, it may be infeasible or impossible to completely randomize the order of the trials within blocks. This often results in a so-called **split-plot design**, as illustrated in the following example.

Example 7.8 A fuel fabrication facility must test the effects of sintering time and temperature on the density of UO_2 pellets. Three levels each of sintering time and temperature are to be investigated. The experiment is conducted for each of three lots of pellets in turn by heating the furnace to one of the temperature levels and sintering three pellets for each of the three sintering times being studied. The process is repeated for each of the other two temperature levels.

The three lots of pellets are blocks in this experiment. If each of the $3 \times 3 = 9$ trials in a block could be run in completely random order, this would be a randomized complete block design; however, this is not possible. For any given lot of UO_2 pellets, the three trials associated with the first temperature are run, then the three trials associated with the second temperature are run, and finally those associated with the last temperature are run. The three temperature levels are referred to as **plots**. Each sintering time within a plot is a subplot or **split-plot**. This terminology is derived from agricultural experiments where each of several levels of a factor is tested on each of several plots of land. The plots are split into subplots to investigate levels of another factor.

Two particular aspects of this design are worth noting. First, the plots define a restriction on randomization within a lot (block) of UO_2 pellets. That is, plots are the levels of a second blocking factor. Second, the plots correspond directly to levels of the temperature factor. Thus, if there are any uncontrolled factors that vary as the temperature levels are changed, the effects of these uncontrolled factors are confounded with the temperature effects. Another way of saying this is that the temperature main effect is completely confounded with the plot effect.

Although there are two restrictions on randomization, it is not completely restricted. Nine pellets are randomly selected from each lot and are randomly assigned to each of the nine trials (time-temperature combinations). The order in which the levels of temperature are applied should be randomized. It may seem more natural to consider the lowest temperature first, followed by the next two higher temperatures in turn. This order, however, may aggravate any existing problems due to the confounding of temperature with plots. Running the temperature levels in random order helps to distribute evenly some of the uncontrolled factors that may contribute to the plot effect. Within a single plot, pellets are to be sintered for three different lengths of time. Here, theory suggests that the three sintering times should be observed at random and not in order. Practically, however, it is easier to observe the three times in order. If the times are fairly long, it may be that an experimenter would put the three pellets into the furnace simultaneously and then remove each one when its time is up. This more practical scheme would not evenly distribute the effects of any

uncontrolled factors that vary along with time. Thus, interpretation of the time effect would require an assumption that the effects of such uncontrolled factors are negligible.

To illustrate the previous discussion of randomization, assume that nine pellets are randomly selected from each lot and are randomly assigned to each of the nine trials within a block (lot of pellets). One possible run order, which employs the restricted randomization discussed previously, is displayed in Table 7.30. With this run order, block (pellet lot) number 2

TABLE 7.30
Example of Restricted Randomized Run
Order for Split-Plot Design

		Blocks (lots of UO ₂ pellets)								
		1			2			3		
Temperature levels		1	2	3	1	2	3	1	2	3
Sintering times										
1		10	17	14	6	2	7	21	26	22
2		11	16	15	4	3	9	20	27	24
3		12	18	13	5	1	8	19	25	23

is run first, followed by blocks 1 and 3. The run orders of plots (temperature levels) within blocks and split-plots (sintering times) within plots are also determined at random.

The model for this design is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \epsilon_{ijk} \quad (7.7)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, c$$

where

y_{ijk} = the observed density of a pellet from lot i for sintering temperature j and time k

μ = overall mean density of pellets in all lots

α_i = effect of lot i on density

β_j = effect of sintering temperature j on density

γ_k = effect of sintering time k on density

ϵ_{ijk} = the effect of random experimental error on y_{ijk}

and $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$, $\beta\gamma_{jk}$, and $\alpha\beta\gamma_{ijk}$ are the associated interaction terms. Because this experiment is not replicated, the experimental error variance σ_e^2 cannot be estimated.

The split-plot design is often used by necessity rather than by choice. Having the plot effect confounded with a treatment effect is usually not a desirable design property. It may be, however, that such a design is the only feasible or cost-effective method of gathering data in a particular situation.

Further discussions of the development and statistical analysis of split-plot designs may be found in most experimental design texts, such as Montgomery (1976), Hicks (1982), and Cochran and Cox (1957).

7.6 NESTED (HIERARCHICAL) DESIGNS

In many nuclear material accounting experimental situations the levels of one factor (say factor A) are not identical over the levels of another factor (say factor B). Then factor A is said to be **nested** in factor B and the resulting design is referred to as a **nested** or **hierarchical design**.

Methods for analysis of experimental data from nested designs are presented in Section 5.3. Some of the experimental design considerations are illustrated in the following example.

Example 7.9 An experiment is being planned where four samples are to be randomly drawn from each of three randomly selected containers of ammonium diuranate (ADU) scrap, with three uranium analyses performed on each sample. The purpose of the experiment is to provide data for estimating variance components for containers, sampling, and chemical analysis. This experimental layout is displayed below.

		ADU Containers											
		1				2				3			
Samples		1	2	3	4	1	2	3	4	1	2	3	4
Analyses													
	1												
	2												
	3												

There are three factors in this experiment: ADU containers, samples, and analyses. Each of these factors is assumed to have random effects. Although four samples are taken from each container, the samples differ from container to container. Thus, the factor "samples" is said to be nested within the factor "containers." Likewise, "analyses" are nested within "samples." Thus, this particular experiment is referred to as a nested (or hierarchical) experiment. Because the factors have random effects, the variance components for these factors are of interest in estimating the contribution of each factor to the total uncertainty in the uranium analysis results.

The model for this problem is given by Equation 5.101, which is repeated here for convenience; that is,

$$y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + \epsilon_{(ij)k} \quad (7.8)$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, n$$

where

y_{ijk} = observed percent uranium from analysis k on sample j from container i

μ = overall mean percent uranium in population of ADU containers

α_i = effect of i^{th} container

$\beta_{(i)j}$ = effect of j^{th} sample within i^{th} container

$\epsilon_{(ij)k}$ = effect of k^{th} analysis within j^{th} sample and i^{th} container

In developing this nested design, the principle of randomization is incorporated both in assigning experimental units to and in determining the run order of the various trials. In this experiment, the experimental units are subsamples of the original samples taken from each container. Because of the three-stage random sampling, it is not essential to randomly assign the subsamples to trials; randomization of the run order suffices for this problem. Note, however, that random assignment of experimental units may be required in other experiments. The run order of all $3 \times 4 \times 3 = 36$ trials should be completely randomized. One such possible randomized run order is shown in Table 7.31.

Suppose that the data displayed in Table 7.32 are obtained by performing the experiment. The nested ANOVA techniques presented in Section 5.3.1 are used to obtain the results displayed in Table 7.33. The F-ratios indicate that the variance components $\sigma_{\text{Containers}}^2$ and $\sigma_{\text{Samples}}^2$ are both greater than zero. The quantity $\sigma_{\text{Analyses}}^2$ represents the variability due to

TABLE 7.31
One Possible Randomized Run Order for Nested Experiment

Samples	ADU containers											
	1				2				3			
	1	2	3	4	1	2	3	4	1	2	3	4
Analyses												
1	21	18	3	11	33	26	17	31	1	13	23	7
2	20	12	36	16	8	6	25	14	34	4	19	32
3	5	35	30	15	10	24	27	2	29	9	28	22

TABLE 7.32
Uranium Analysis Results in Percent Uranium
for Nested ADU Experiment

Samples	ADU containers											
	1				2				3			
	1	2	3	4	1	2	3	4	1	2	3	4
Analyses												
1	44.3	33.8	42.3	38.2	24.1	27.3	22.7	25.6	79.8	76.4	77.0	75.4
2	44.2	33.0	42.7	37.9	24.2	26.7	22.9	25.9	80.2	76.2	77.4	75.8
3	44.4	33.3	42.5	38.3	24.5	26.9	22.8	26.0	79.7	76.8	77.3	75.9

TABLE 7.33
ANOVA Table for Two-Stage Nested ADU Experiment

Source	df	SS	MS	F
Containers	2	17,519.444	8759.722	$\frac{8759.722}{30.485} = 287.35$
Samples within containers	9	274.365	30.485	$\frac{30.485}{0.060} = 508.08$
Analyses within samples	24	1.447	0.060	
Total	35	17,795.256		

individual analyses and experimental errors and cannot be statistically tested.

The purpose of this experiment is to estimate the variance components due to sampling containers from the population, sampling material from within containers, and analysis of the sampled material. By applying the variance component estimation method discussed in Section 5.3.1.1, the expected mean squares are derived and used to calculate the desired variance components. The expected mean squares (Table 5.4) are summarized as follows:

Source	E[MS]
Containers	$\sigma_{\text{Analyses}}^2 + 3\sigma_{\text{Samples}}^2 + 12\sigma_{\text{Containers}}^2$
Samples	$\sigma_{\text{Analyses}}^2 + 3\sigma_{\text{Samples}}^2$
Analyses	$\sigma_{\text{Analyses}}^2$

By equating these expected mean squares to the calculated mean squares in Table 7.33, the variance component estimates are computed as follows:

$$\hat{\sigma}_{\text{Analyses}}^2 = \text{MS}_{\text{Analyses}} = 0.060$$

$$\begin{aligned} \hat{\sigma}_{\text{Samples}}^2 &= \frac{\text{MS}_{\text{Samples}} - \text{MS}_{\text{Analyses}}}{3} \\ &= \frac{30.485 - 0.060}{3} = 10.142 \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}_{\text{Containers}}^2 &= \frac{\text{MS}_{\text{Containers}} - \text{MS}_{\text{Samples}}}{12} \\ &= \frac{8759.722 - 30.485}{12} = 727.436 \end{aligned}$$

An estimate of the total variance of y_{ijk} is given by

$$\widehat{\text{Var}}(y_{ijk}) = 727.436 + 10.142 + 0.060 = 737.638$$

The estimated standard deviations for analyses, samples, and containers are computed as the square roots of the variance component estimates;

they are 0.245, 3.185, and 26.971% uranium, respectively. The estimated standard deviation of y_{ijk} is

$$\sqrt{\hat{\text{Var}}(y_{ijk})} = \sqrt{737.638} = 27.159 \% \text{ uranium}$$

An experiment with nested factors might involve only factors with random effects or it might include factors with fixed effects as well. The calculation of ANOVA tables and the estimation of variance components and means for such designs are discussed in Section 5.3. Additional discussion of the development and statistical analysis of nested designs is found in Montgomery (1976) or Hicks (1982).

REFERENCES

- Box, G. E. P., and J. S. Hunter, 1961a,b, The 2^{k-p} Fractional Factorial Designs, Parts I and II, *Technometrics*, 3: 311-352; 449-458.
- , W. G. Hunter, and J. S. Hunter, 1978, *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.
- Cochran, W. G., and G. M. Cox, 1957, *Experimental Designs*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Connor, W. S., and M. Zelen, 1959, *Fractional Factorial Experimental Designs for Factors at Three Levels*, Applied Mathematics Series No. 54, National Bureau of Standards, Washington, D. C.
- Davies, O. L., 1956, *Design and Analysis of Industrial Experiments*, Hafner Publishing Co., New York.
- Fisher, R. A., and F. Yates, 1953, *Statistical Tables for Biological, Agricultural and Medical Research*, 4th Ed., Oliver and Boyd, Edinburgh.
- Hicks, C. R., 1982, *Fundamental Concepts in the Design of Experiments*, 3rd Ed., Holt, Rinehart and Winston, New York.
- Kemthorne, O., 1973, *The Design and Analysis of Experiments*, R. E. Krieger Publishing Co., Huntington, New York.
- Montgomery, D. C., 1976, *Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York.
- Scheffé, H., 1959, *The Analysis of Variance*, John Wiley & Sons, Inc., New York.
- Snedecor, G. W., and W. G. Cochran, 1980, *Statistical Methods*, 7th Ed., Iowa State University Press, Ames, Iowa.

ADDITIONAL READING

- Anderson, V. L., and R. A. McLean, 1974, *Design of Experiments, A Realistic Approach*, Marcel Dekker, Inc., New York.
- John, P. W. M., 1971, *Statistical Design and Analysis of Experiments*, Macmillan, New York.
- Kirk, R. E., 1982, *Experimental Design*, 2nd Ed., Wadsworth, Inc., Belmont, California.
- McLean, R. A., and V. L. Anderson, 1984, *Applied Factorial and Fractional Designs*, Marcel Dekker, Inc., New York.

- Ostle, B., and R. W. Mensing, 1979, *Statistics in Research*, 3rd. Ed., Second Printing, Iowa State University Press, Ames, Iowa.
- Raktoc, B. L., A. Hedayat, and W. T. Federer, 1981, *Factorial Designs*, John Wiley & Sons, Inc., New York.
- Winer, B. J., 1971, *Statistical Principles in Experimental Design*, 2nd Ed., McGraw-Hill, Inc., New York.

CHAPTER 8

Statistical Sampling of Finite Populations

8.0 INTRODUCTION

Two common activities in nuclear material accounting are the determination of the special nuclear material (SNM) content of an inventory and the verification of prior location and content information for various containers (items) within an inventory. Often these activities are performed by physically inspecting all containers or locations of SNM within a facility. This complete census approach can be time-consuming and expensive. In some cases, it may be possible to satisfactorily estimate the desired information by inspecting only a small portion (sample) of all containers or locations of SNM in a facility by using statistical sampling techniques.

Statistical sampling also plays an important role in other nuclear material accounting problems, for example:

1. Isotope or element factors are often provided by lot (or batch) for a group of SNM containers. These are calculated by estimating the factor for each of several containers sampled from the lot and then using the mean of the sample values to estimate the true mean factor for the lot.

2. In the production of fuel pellets, the fuel producer must control the uranium content of the product. Because destructive analysis is required to verify the uranium content of the pellets, it is not possible to measure every pellet in a particular batch. A sample of pellets is taken from the batch and analyzed, and a hypothesis is tested about the mean uranium content of all pellets in the batch.

In the above discussion and examples, the populations being sampled are finite. Most of the statistical methods considered in the previous chapters assume that a random sample is taken from an infinitely large population, or a population large enough to be treated as infinite. In nuclear material inventory control, however, the population of interest is often a finite number of containers that are being stored at various locations within a facility. The statistical sampling techniques presented in this chapter are applicable when the population under investigation is finite.

The concept of a sample is that a portion of all possible objects or observations (elements) is considered. The set of all possible objects or observations is referred to as the **population**. A **sample** is a subset of the

objects or observations selected from the population. There are a number of statistical sampling techniques that can be used to select a sample from the population. A **sampling plan** is a series of steps required to select a sample, including problem formulation, selection of one or more basic sampling techniques, specification of precision requirements, calculation of required sample size, and development of a protocol for selecting the sample from the population.

Sections 8.1 and 8.2 present a brief introduction to six basic sampling techniques and discuss the steps required to develop a sampling plan. Four of the basic sampling techniques, with some variations and extensions, are presented in greater detail in Section 8.3. These sections explain how to develop sampling plans that are based on some of the basic sampling techniques. They also provide formulas for estimating population means and totals and the standard errors of the estimates and formulas for determining required sample sizes. Sampling for verification purposes in nuclear material management is presented with examples in Chapter 17.

The statistical sampling literature has developed a special notation for use in equations. In many cases, this notation differs from the notation used for similar concepts in other chapters of this text. One major difference is the use of upper-case letters to denote some population parameters and lower-case letters to denote the corresponding sample estimators *and* estimates. For example, the population mean is denoted by \bar{X} and the sample mean is denoted by \bar{x} . Because the notation varies within the statistical sampling literature, it was necessary to choose one convention and then use it consistently. The notation used by Levy and Lemeshow (1980) was chosen for this chapter. Unfortunately, this may cause some confusion when the reader refers to other chapters of this text. With this caveat in mind, however, the material presented in this chapter should provide valuable guidance in developing and implementing sampling plans and in computing parameter estimates from the resulting data.

The sampling and estimation methods presented in Section 8.3 are from the classical sampling theory literature, and most of them were originally developed for survey sampling of human populations. The theory underlying each of these methods is based on the assumption that the value of the variable or characteristic of interest can be measured or otherwise determined without error for each sampled object. Unfortunately, this assumption is not valid for the types of nuclear material management applications mentioned above where, for example, the contents (or some characteristic of the contents) must be determined for containers that have been sampled from those on inventory. In such applications, the problem of selecting a sample of containers from those on inventory can usually be solved by applying one or more of the finite population sampling techniques presented in Section 8.3. However, the steps of subsampling, measurement, and/or analysis of the material within each container are subject to errors that should not be ignored.

Consequently, in many applications, the overall sampling and analysis is an application of multistage sampling, analogous to the nested or hierarchical experimental designs discussed in Sections 5.3 and 7.6, except that the first stage of sampling (containers) is from a finite population, and the remaining stages of sampling are from populations that are conceptually infinite (i.e., all possible subsamples of material from each container and all possible analyses of each material sample). Thus, all stages of sampling must be carefully designed using either the methods presented in this chapter or those in Chapter 7, and the variances from the error sources at each stage must be taken into account in interpreting any parameter estimates or inferences that are based on the sampling data.

With the exception of two-stage cluster sampling (Section 8.3.2.2), no multistage sampling techniques are presented in this chapter. It should be noted, however, that the methods presented in Section 8.3 and those in Chapter 7 can be used as "building blocks" in developing an overall (multistage) sampling and analysis plan. That is, a different sampling technique might be used at each stage of sampling. For most practitioners, the only difficulty with multistage sampling plans is that estimation formulas are not always readily available. This should not preclude the use of multistage sampling, because familiarity with the basic sampling techniques is sufficient to formulate a sampling plan that is tailored to a specific application. Cochran (1977), Levy and Lemeshow (1980), Yamane (1967), and other books on sampling provide discussions and estimation formulas for some multistage sampling plans. For cases not found in a book or other reference, estimation formulas and a protocol for implementation of the plan can be derived by a statistician with training in sampling theory.

8.1 BASIC SAMPLING TECHNIQUES

The word **sampling** has two meanings in nuclear material management. It is used in a statistical context to refer to the selection of a subset from a population, such as inventory items, record entries, or containers that are to be (or have been) inspected. It is also used in a physical context to describe a small portion of bulk material that is analyzed in a laboratory or otherwise measured or tested. Thus, for example, a sample of uranium oxide drums is selected from those on inventory and a sample of powder from each is sent to the laboratory for analysis.

Random sampling is a method or protocol for taking a sample from a group of distinct items. A random sample is not a haphazardly selected sample, but is chosen by a prespecified procedure that sometimes uses random numbers to determine which items will be selected. A haphazard or seemingly purposeless choice is generally insufficient to guarantee randomness because of subjective, and usually unconscious, biases inherent in personal choice. Samples are random only if drawn according to the requirements of the definitions given below. No matter which method is used to

select a sample, an unambiguous statement of that method is a necessary part of any subsequent interpretation of the data derived from the sample.

In principle, a portion of the contents of an item or container is also chosen randomly from all possible similar portions within the container. In practice, however, this is not often workable because of the logistics of handling large containers. For example, consider drawing a 5-g sample from a drum of uranium oxide powder. In principle, the drum would be completely divided into 5-g sections, one would be chosen randomly, and then by some mechanical means that one section would be removed from the drum. In practice, however, the contents of the container must be made as nearly homogeneous as possible so that a sample taken from any location is equivalent to a random sample. Thus, the population of all possible 5-g samples that could be selected from this container is a conceptual population and is probably best treated as an infinite population.

The following brief descriptions provide introductions to six basic techniques for sampling from finite populations. Four of these techniques (simple random sampling, stratified random sampling, cluster sampling, and probability proportional to size sampling) are presented in greater detail and with estimation formulas in Section 8.3.

1. Simple Random Sampling: A procedure in which all possible samples of some fixed size have an equal chance of being chosen. Elements may be chosen from the population **with replacement** (an element may be chosen more than once) or **without replacement** (an element may be chosen only once).

2. Systematic Random Sampling: A procedure in which an element is randomly selected from among the first k elements of the population, and then every subsequent k^{th} element following the first is included in the sample. For example, if $k = 10$, and the third element is initially selected, then the 3rd, 13th, 23rd, 33rd, . . . , elements are in the sample.

3. Stratified Random Sampling: A procedure in which a population is divided into mutually exclusive and exhaustive strata and a sample is randomly selected from each stratum. The strata are formed so that the elements within each strata are as homogeneous as possible with respect to the characteristic(s) being measured, but the variability between strata is maximized; that is, the strata are chosen to be as different as possible with respect to the characteristic(s) being measured. Then by taking a random sample (of at least one element) from each stratum, this technique (when the stratification is successful) tends to give better coverage of the population than simple random sampling.

4. Cluster Sampling: A procedure in which groups (clusters) of population elements are selected instead of individual elements. Clusters are often chosen by simple random sampling or by probability proportional to size sampling (defined below). Multistage cluster sampling involves sampling of subclusters or individual elements from within clusters chosen at the previous stage(s) of sampling.

5. Probability Proportional to Size (PPS) Sampling: A procedure in which population elements or clusters of elements are randomly chosen with probabilities of selection proportional to size. "Size" is usually a measurable characteristic associated with the relative importance of a population element, for example, an amount of SNM, or a count of the number of elements within a cluster.

6. Sequential Sampling: A procedure in which a decision rule is applied sequentially to sample elements. The concept of sequential sampling and hypothesis testing is presented in Section 4.6 along with examples. Sequential sampling may be conducted with any of the sampling techniques described above. It is useful because, on the average, it requires smaller sample sizes than nonsequential sampling techniques. There are three basic types of sequential sampling:

- a. **Simple Sequential Sampling:** A test of significance is performed after each sample element is selected.
- b. **Curtailed Sequential Sampling:** This is simple sequential sampling where a maximum sample size is specified.
- c. **Multistage Sequential Sampling:** Sample elements are selected in small groups or stages, and the groups are sequentially selected and tested. The number of stages may be curtailed.

Simple sequential sampling is advantageous when samples are difficult or expensive to collect or analyze. Multistage sequential sampling is typically used when sample analysis is time consuming but samples can be analyzed together in batches.

As explained in Section 8.0, each sampling technique discussed above should be considered as a basic or "building block" technique. The process of designing a sampling plan takes into consideration the strengths and weaknesses of the basic sampling techniques and then combines them into a plan best suited to the problem and its underlying constraints. Some of the fundamental strengths and weaknesses of the above six basic sampling techniques are given in Table 8.1.

Several of the comments in Table 8.1 concern the estimation of parameters based on data collected by a sampling technique. These comments refer to situations where estimates of parameters, such as a mean and variance, are required. Comments about the feasibility and the effort required for a specific sampling technique are general and apply to most applications.

8.2 DEVELOPING A SAMPLING PLAN

The major steps in developing a sampling plan are outlined below with comments on their interrelationships.

1. Formulate the problem. A nuclear material accounting problem should be stated so that the data to be collected are meaningful and are directly applicable to the investigation and solution of the problem.

TABLE 8.1
Strengths and Weaknesses of Basic Sampling Techniques

Sampling technique	Strengths	Weakness
1. Simple random	<p>Selection of sample elements easy to perform</p> <p>Unbiased estimates of mean and variance</p>	<p>Requires all population elements to be identified and labeled prior to sampling</p> <p>Can be time-consuming/expensive to locate selected elements within the population</p> <p>May not adequately represent important subgroups that comprise a small portion of the population</p>
2. Systematic	<p>Where applicable, the selection and gathering of sample elements is easy</p> <p>Does not require knowledge of total number of population elements</p>	<p>May require ordering of population elements</p> <p>Simplifying assumptions required to estimate parameters</p> <p>May produce biased parameter estimates</p>
3. Stratified	<p>Includes at least one element from each stratum; thus, if stratification is successful, better coverage of the population is achieved than with other techniques</p> <p>Allows separate decisions or parameter estimates for each stratum as well as for total population. Note, however, that sample sizes calculated to provide a certain precision for inferences about the population do not provide the same precision for inferences within a stratum</p>	<p>Elements within strata should be relatively homogeneous</p> <p>Time-consuming to identify sample elements</p> <p>More sample elements may be required than for other techniques</p>

Increased precision (smaller standard errors)
under certain conditions

Stratification can sometimes be applied after
the sample has been collected by another
sampling technique

4. Cluster	Does not require a list of all population elements to select sample units May save time/cost Multistage clustering is advantageous for large populations with naturally occurring hierarchical structure	May ignore clusters of interest May underestimate standard errors for some parameters if the elements within the clusters are homogeneous
5. Probability proportional to size	Allows sampling according to some measure of size or importance of each element to ensure coverage of the areas of interest in the population	Highly dependent upon measure of size or importance used
6. Sequential	On the average, smaller sample sizes are required than with non-sequential techniques	Not applicable in all situations Requires involved calculations after acquiring each sample element for some applications The possibility exists that sequential sampling may yield sample sizes larger than required by non-sequential techniques The average sample size required to reach a decision depends upon the decision rule (statistical test) used

2. Identify the population and its characteristics. The definition of the population of interest follows from a formulation of the problem. Characteristics of the population to be studied (such as number and variety of SNM containers, storage method, range of SNM content per container, and so on) should be known. A list of population elements should be available.

3. Specify the conditions of the problem. If the problem is to detect a loss in the overall inventory, the values desired for the loss detection goal, probability of detection, and probability of a false alarm (concluding there is a loss when, in fact, there is not) should be specified. For the estimation of a quantity, such as an isotope factor or the total quantity of SNM in an inventory, the required closeness of the estimate to the true value and the desired probability of this closeness being achieved must be stated.

4. Specify the measurement model and assumptions. A statistical model and its assumptions are specified on the basis of measurement techniques to be used. An estimate of the population variance (or variance components in the case of multistage sampling plans) is required to calculate the required sample size (see Chapter 14 for a discussion of measurement models).

5. Identify the appropriate statistical methods. Identify the statistical tests or estimation procedures and basic sampling technique(s) required to solve the problem formulated in Step 1. The formulas for calculation of required sample size (or sizes) depend on the basic sampling technique(s) and the test or estimation method to be utilized.

6. Perform preliminary preparations if required. The population characteristics may indicate that stratified or cluster sampling techniques should be employed. If so, the strata or clusters within the population must be identified. A list† and count of the elements in each stratum or cluster should be developed.

7. Calculate the required sample size. The sampling distribution associated with the statistical method identified in Step 5 provides the basis for a sample size formula. A sample size is obtained by substituting the problem requirement values from Step 3 into the appropriate sample size formula. For sequential sampling, the sample size‡ is dependent upon intermediate results.

8. Allocate the sample size. Once sample size (or sizes) is determined, a protocol is developed whereby the sampling techniques (discussed in Section 8.1) are used to allocate the sample size among elements, strata, or clusters of the population.

†For cluster sampling, only a list of the population elements within sampled clusters is required. A complete listing over all clusters is not essential.

‡Expected sample sizes can be computed for most sequential techniques.

9. **Select the sample.** The specific units (containers, clusters, or batches of containers) are selected according to the allocation and protocol (sampling plan) developed in Step 8.

The sampling plans considered in Section 8.3 are described primarily in terms of Steps 5 through 8, although the other steps are discussed as necessary. Also, for most applications of finite population sampling in nuclear material control, the SNM will be in drums, cans, cylinders, buckets, or other such containers. Therefore, in Section 8.3, such items in a population are sometimes referred to as containers.

8.3 SAMPLING PLANS FOR ESTIMATING MEANS AND TOTALS

Many nuclear material accounting activities have a goal of estimating a population mean or total from a sample. For example, an estimate of the total amount of SNM in a facility is required at inventory time. An estimate of the mean element or isotope factor may be needed for a batch of UO_2 powder held in many containers. Because of sampling and measurement errors, standard error estimates and confidence intervals are needed along with the point estimate of a mean or total. (The standard error is defined in Chapter 3 as the standard deviation of a parameter estimator.)

The required sample size for estimating a population mean or total is indirectly determined by specifying how close the estimate must be to the true value and the probability that this closeness will be achieved. Let x be the sample estimate of a mean or total and X be the true but unknown mean or total. In practice, closeness can be specified by placing a limit ϵ on the relative difference between x and X ; that is,

$$\frac{|x - X|}{X} \leq \epsilon \quad (8.1)$$

Because of sampling and measurement variability, Equation 8.1 cannot always be satisfied, but a probability $1 - \gamma$ can be specified with which it is expected to be satisfied; that is

$$\Pr\left(\frac{|x - X|}{X} \leq \epsilon\right) = 1 - \gamma \quad (8.2)$$

The sample estimator x is a function of either the sum or a weighted sum of the measured values for the sampled objects. Thus, in many applications, x can be assumed to be approximately normally distributed (see Section 2.6.2.1 for a discussion of this concept), so that Equation 8.2 can

be written in terms of the cumulative distribution function (cdf) of the standard normal distribution. This requires a value for the population variance of the characteristic of interest. In multistage sampling applications, values of the variance components for each stage of sampling are required. In practice, however, the population variance is usually unknown, so an estimate is substituted based on prior experience or a preliminary sample. For multistage sampling, estimates of the variance components are substituted. This results in an expression that is a function of the sample size (or sizes in multistage sampling), the population variance (or variance components in multistage sampling), and the precision requirements ϵ and γ .

For the single-stage sampling applications of the techniques presented in this section, required sample size formulas are derived from Equation 8.2. The derivations do not include cost considerations, and they follow the same basic logic as the derivation of Equation 3.28 in Section 3.5.2. As mentioned in Section 8.0, however, the sampling techniques presented in this section were developed under the assumption that the true value of the variable or characteristic of interest can be measured or otherwise determined without error for each sampled object. Because this assumption may not be valid in most nuclear material management applications, the sample size formulas presented in this section should be applied with caution.

In fact, unless the variance components for subsampling, measurement, analytical, and other errors are negligible compared to the variance in the true values among the objects in the population, application of the sample size formulas presented in this section could underestimate the required sample size and lead to an inadequate sampling plan. When the variance components for these sources of error are not negligible, a multistage sampling plan may be appropriate. In such cases, sample size requirements for the various stages of sampling can be derived from Equation 8.2; the methods involved are trial-and-error and/or iterative in nature, and they are not presented in this chapter.

To present some of the basic finite population sampling techniques and illustrate their implementation, each technique (except two-stage cluster sampling) is presented as a single-stage sampling technique, where it is assumed that subsampling, measurement, analytical, and other errors are negligible. If this assumption is not valid, the formulas for estimating a population mean or total can still be used, but the formulas for the standard error estimates will tend to underestimate the true standard error of the estimator of a mean or total.

It is very difficult to find realistic nuclear material management examples where this assumption is strictly valid for single-stage sampling of a finite population. Thus, rather than presenting contrived examples that satisfy this assumption, some realistic examples are presented that probably do not satisfy the assumption. However, with this caveat in mind, the examples serve to illustrate implementation of the techniques.

8.3.1 Simple Random Sampling Plans

Simple random sampling is usually easy to perform and yields simple formulas for the estimator of a population mean or total, its standard error, and the sample size required to meet specified precision requirements. To select a simple random sample without replacement, assign each element in the population a number from 1 to N , where N is the number of elements in the population. Next, select n of these numbers by some random process such as a table of random numbers, or a calculator or computer with a random-number generator (see Sections 2.5.1 and 11.3.1). Sampling without replacement means that each element identification number chosen to be in the sample can appear in the sample only once. Once the n distinct numbers are chosen, the population elements corresponding to these numbers are taken from the population as the sample.

The estimators for a population mean and total and their estimated standard errors under simple random sampling (without replacement) are as follows:

Population Mean, \bar{X}

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (8.3)$$

$$\hat{SE}(\bar{x}) = \left(\frac{N-n}{nN} \right)^{1/2} \hat{\sigma}_x \quad (8.4)$$

Population Total, X'

$$x' = N\bar{x} \quad (8.5)$$

$$\hat{SE}(x') = N\hat{SE}(\bar{x}) \quad (8.6)$$

where

n = number of elements in the sample

N = number of elements in the population

\bar{x} = estimator for the population mean

x' = estimator for the population total

x_i = value of the variable of interest for the i^{th} sample element

\hat{SE} = estimated standard error

$$\hat{\sigma}_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \text{sample standard deviation}$$

Approximate two-sided $100(1 - \alpha)\%$ confidence intervals for the population mean and total are given, respectively, by

Population Mean, \bar{X}

$$\bar{x} \pm z_{1-\alpha/2} \hat{SE}(\bar{x}) \quad (8.7)$$

Population Total, X'

$$x' \pm z_{1-\alpha/2} \hat{SE}(x') \quad (8.8)$$

where $z_{1-\alpha/2}$ is obtained from Table A3. Additional comments concerning these confidence intervals and the validity of the approximation are given in Section 8.3.5.

The sample size necessary to achieve prespecified precision requirements γ and ϵ for both mean and total estimators is given by

$$n = \frac{(z_{1-\gamma/2})^2 N (\hat{CV})^2}{(z_{1-\gamma/2})^2 (\hat{CV})^2 + (N - 1)\epsilon^2} \quad (8.9)$$

where

$\hat{CV} = \frac{\hat{\sigma}_x}{\bar{x}}$ = an estimate of CV, the population coefficient of variation, from previous knowledge or the most recent sample estimate

ϵ = maximum relative difference limit on $\frac{|x' - X'|}{X'}$ or $\frac{|\bar{x} - \bar{X}|}{\bar{X}}$

γ = probability that the absolute relative difference between the sample estimate and population value is greater than ϵ

Equation 8.9 is from Levy and Lemeshow (1980), and the details of its development can be found in the literature review by Piepel and Brouns (1982). This formula provides a required sample size for simple random sampling without replacement.

Equations 8.3 through 8.6 are based on simple random sampling without replacement. In nuclear material accounting applications, there is generally no advantage to sampling with replacement when simple random sampling is used. Thus, sampling with replacement is not discussed here.

Example 8.1 A fuel fabrication facility has 1000 trays of high-enriched (HEU) fuel pellets, with each tray containing between 1.4 and 1.6 kg of ^{235}U in HEU. An estimate of the total ^{235}U content of the trays of pellets is required, and it must be within 10 kg of the unknown true population value with probability 0.95. In the last complete inventory, a similar group of pellet trays had a mean ^{235}U content of 1.5 kg per tray and a standard deviation of 0.058 kg.

In this problem, the population to be sampled is the 1000 trays of pellets. Assume that the average amount of ^{235}U per tray is still approximately 1.5 kg, so that the total amount of ^{235}U in the 1000 trays is approximately 1500. For the estimate to be within 10 kg of the population value, Equation 8.1 gives $\epsilon \cong 10/1500 = 0.0067$. For probability 0.95, $\gamma = 0.05$, and the quantity $z_{(1-\gamma/2)} = z_{0.975} = 1.96$ is obtained from Table A3. The coefficient of variation from the last complete inventory is $\text{CV} = (0.058)/(1.5) = 0.0387$. By substituting these values of ϵ , γ , and CV into Equation 8.9, the required sample size is

$$n = \frac{(1.96)^2(1000)(0.0387)^2}{(1.96)^2(0.0387)^2 + (999)(0.0067)^2} = 113.71 \cong 114$$

A sample of 114 trays was selected by simple random sampling without replacement and the ^{235}U content of each was measured, yielding

$$\sum_{i=1}^{114} x_i = 170.56$$

and

$$\hat{\sigma}_x = 0.059$$

By substituting these values and those of N and n into Equations 8.5, 8.6, and 8.8, an estimate of the population total and its estimated standard error are computed as

$$x' = N\bar{x} = 1000 \left(\frac{170.56}{114} \right) = 1496.14 \text{ kg}$$

and

$$\hat{S}\hat{E}(x') = N \hat{S}\hat{E}(\bar{x}) = 1000 \left[\frac{1000 - 114}{(114)(1000)} \right]^{1/2} (0.059) = 5.20$$

An approximate 95% confidence interval for X' , the total ^{235}U content of all trays of pellets in the population, is computed from Equation 8.8, which gives

$$1496.14 \pm (1.96)(5.20)$$

or

$$(1485.95, 1506.33)$$

Thus, we are 95% confident that the total ^{235}U content of this population is between 1485.95 and 1506.33 kg. Note that the half-width of this confidence interval is $1.96 (5.20) = 10.19$ kg, which is in very close agreement with the precision requirement of 10 kg.

Example 8.2 A batch of UO_2 powder is stored in 30 cans, and an estimate of the mean uranium factor for the batch is needed. An estimate within 1% of the true batch mean is required with probability 0.98. Past information suggests that $\text{CV} = 0.0195$. How many cans should be sampled?

Equation 8.9 provides the sample size necessary to meet the precision requirements ($\gamma = 0.02$ and $\epsilon = 0.01$), and yields the value

$$n = \frac{(2.33)^2 (30) (0.0195)^2}{(2.33)^2 (0.0195)^2 + (29)(0.01)^2} = 12.48$$

Thus, 13 of the 30 cans must be selected by simple random sampling without replacement in order to meet the specified precision requirements.

8.3.1.1 Limitations of Finite Population Sampling Methods

It was emphasized previously that the finite population sampling techniques presented in this chapter were developed under the assumption that the value of the variable or characteristic of interest can be determined

without error for each sampled object. It was also emphasized that this assumption is rarely, if ever, valid in nuclear material control applications. Therefore, the techniques presented in this chapter are rarely applicable as "stand-alone," single-stage sampling methods. Instead, they are applicable as building block techniques in an overall multistage sampling and analysis plan. At this point, it is instructive to discuss the implications of applying one of these techniques, by itself, in the presence of subsampling, measurement, and/or analytical errors.

Consider a situation like the one described in Example 8.2. Suppose that n cans of UO_2 powder are selected by simple random sampling from the $N = 30$ cans on inventory. If the true percent uranium could be determined without error for each can in the sample, then the sample mean \bar{x} , computed from Equation 8.3, would have variance given by

$$\text{Var}_1(\bar{x}) = \left(\frac{N - n}{N} \right) \frac{\sigma_{\text{Cans}}^2}{n}$$

where σ_{Cans}^2 is the variance of the true percent uranium values among the $N = 30$ cans on inventory. Note that $\text{Var}_1(\bar{x})$ approaches zero as n approaches N . That is, if all N true percent uranium values are in the sample, then the sample mean is the population mean, and therefore has zero variance. In other words, if $n = N$, the population mean would be known without error.

In practice, however, the true percent uranium could not be determined without error for each can in the sample. Suppose that a powder sample is taken from each of the n cans in the sample and that an analysis is performed on each powder sample. Then x_i is the observed percent uranium value for the i^{th} can. If \bar{x} is computed from Equation 8.3, the variance of \bar{x} is given by

$$\text{Var}_2(\bar{x}) = \left(\frac{N - n}{N} \right) \frac{\sigma_{\text{Cans}}^2}{n} + \frac{\sigma_{\text{Samp}}^2}{n} + \frac{\sigma_{\text{Anal}}^2}{n}$$

where σ_{Samp}^2 and σ_{Anal}^2 are the variance components for subsampling and analysis. Because by definition variance components are non-negative, $\text{Var}_2(\bar{x})$ is a larger quantity than $\text{Var}_1(\bar{x})$, where the subsampling and analytical errors were assumed not to exist. If n is set equal to N , note that only the first term in $\text{Var}_2(\bar{x})$ is zeroed out, leaving a nonzero variance. This conflicts with the theory of finite population sampling.

For this example, the square of the true standard error of \bar{x} , $\text{SE}^2(\bar{x})$, is equal to $\text{Var}_2(\bar{x})$. However, if Equation 8.4 is applied to compute an

estimate of $SE(\bar{x})$, the square of this estimate actually estimates the quantity

$$E[\hat{SE}^2(\bar{x})] = \left(\frac{N-n}{N} \right) \left[\frac{\sigma_{\text{Cans}}^2}{n} + \frac{\sigma_{\text{Samp}}^2}{n} + \frac{\sigma_{\text{Anal}}^2}{n} \right]$$

Thus, $\hat{SE}^2(\bar{x})$ underestimates $SE^2(\bar{x}) = \text{Var}_2(\bar{x})$ because it incorrectly underestimates, by a factor of $(N-n)/N$, the contribution of the variance components for subsampling and analysis (e.g., for $N = 30$ and $n = 13$, $(N-n)/N = 17/30 = 0.567$).

This exercise demonstrates that application of a finite population sampling method when the assumptions are not valid (i.e., ignoring the subsampling and analytical error variances in this example) can lead to underestimation of the standard error of the sample mean (or total). The degree or extent of underestimation depends upon the true values of the variance components that are inherently being assumed to be negligible. If in fact they are very small, then the underestimation of the standard error may be of little practical concern. However, this should be investigated and demonstrated before it is assumed to be true.

8.3.2 Cluster Sampling Plans

Clusters of containers occur in nuclear material facilities due to physical storage methods (containers are often stored on shelves, racks, or other devices which naturally group the containers into clusters). Records for each container usually identify only the cluster (room, shelf, etc.) where it is located and not the position of the container within the cluster. Thus, treating clusters as the population elements to be selected according to a sampling plan, and then sampling all containers from each sampled cluster may provide savings in the time and cost required to locate individual containers. See Harkness (1977) for a detailed discussion of container location effort.

There are many types of cluster sampling. The technique mentioned above, where all containers in a selected cluster are sampled, is called one-stage cluster sampling. With two-stage cluster sampling, only a portion of the containers within each selected cluster are sampled. For both one- and two-stage cluster sampling, any of the other sampling techniques can be used to select the clusters or containers within clusters. For example, simple random sampling or PPS sampling is often used to select clusters in one-stage cluster sampling (or the first stage of two-stage cluster sampling). These same sampling techniques are also commonly used for selecting containers from clusters in the second stage of two-stage cluster sampling. One- and two-stage cluster sampling plans based on simple random sampling are discussed and variations of them are considered.

8.3.2.1 Simple One-Stage Cluster Sampling

With simple one-stage cluster sampling, clusters are selected from the population of clusters by simple random sampling. The estimators for a population mean and total and their estimated standard errors are as follows:

Population Mean, \bar{X}

$$\bar{x} = \frac{M \sum_{i=1}^m x'_i}{mN} \quad (8.10)$$

$$\hat{SE}(\bar{x}) = \left(\frac{M}{N\sqrt{m}} \right) \left(\frac{M-m}{M-1} \right)^{1/2} \hat{\sigma}_{clu} \quad (8.11)$$

Population Total, X'

$$x' = N\bar{x} \quad (8.12)$$

$$\hat{SE}(x') = N \hat{SE}(\bar{x}) \quad (8.13)$$

where

\bar{x} = estimator for the population mean

x' = estimator for the population total

m = number of clusters in the sample

M = number of clusters in the population

$x'_i = \sum_{j=1}^{N_i} x_{ij}$ = total value for the i^{th} sample cluster

x_{ij} = value of the variable of interest for the j^{th} container in the i^{th} cluster

N_i = number of containers in cluster i

$N = \sum_{i=1}^M N_i$ = total number of containers in the population

$$\hat{\sigma}_{clu} = \left[\frac{\sum_{i=1}^m (x'_i - \bar{x}_{clu})^2}{m-1} \right]^{1/2} \left(\frac{M-1}{M} \right)^{1/2} = \text{the estimated standard deviation of the cluster totals}$$

$$\bar{x}_{clu} = \sum_{i=1}^m x'_i / m = \text{sample average of the cluster totals}$$

Simple one-stage cluster sampling does not require that the number of containers within each cluster be the same. In fact, the total number of containers sampled, given by

$$n = \sum_{i=1}^m N_i$$

is a random variable and depends upon which clusters are chosen. This scheme gives equal weight to each cluster, regardless of the number N_i of containers within each cluster. Other schemes can be employed that choose clusters with unequal weights; typically the weights are proportional to some measure of size (such as the number of containers per cluster). Probability proportional to size (PPS) sampling plans for containers or clusters of containers are discussed in Section 8.3.3.

With simple one-stage cluster sampling, the number of clusters necessary to achieve prespecified precision requirements γ and ϵ for estimates of the mean and total is given by

$$m = \frac{(z_{1-\gamma/2})^2 M (\hat{C}V_{clu})^2}{(z_{1-\gamma/2})^2 (\hat{C}V_{clu})^2 + (M - 1)\epsilon^2} \quad (8.14)$$

where m and $z_{z-\alpha/2}$ are previously defined and where

$CV_{clu} = \frac{\sigma_{clu}}{\bar{X}_{clu}}$ = the population coefficient of variation among cluster totals ($\hat{C}V_{clu}$ is an estimate from previous knowledge or a sample estimate)

$\sigma_{clu}^2 = \frac{\sum_{i=1}^M (x_i' - \bar{X}_{clu})^2}{M}$ = population variance among all cluster totals

$x_i' = \sum_{j=1}^{N_i} x_{ij}$ = population total for the i^{th} cluster

x_{ij} = value of the variable of interest for the j^{th} container in the i^{th} cluster

$\bar{X}_{clu} = \sum_{i=1}^M x_i' / M$ = true mean of the cluster totals

ϵ = maximum relative difference limit on $\frac{|x' - X'|}{X'}$ or $\frac{|\bar{x} - \bar{X}|}{\bar{X}}$

γ = probability that the absolute relative difference between the sample estimate and population value is greater than ϵ .

Example 8.3 A fuel fabrication facility inventory of low-enriched uranium (LEU) is in 600 containers of UO_2 powder. The containers are stored on 20 moveable racks that hold 30 containers each. Each container holds 12 to 13 kg ^{235}U . Summary information for a complete census of the racks from the previous production run is given in Table 8.2.

TABLE 8.2
Measured ^{235}U Content of 20 Racks of
 UO_2 Containers from a Previous Production Run

Rack	kg ^{235}U	Rack	kg ^{235}U
1	376.851	11	374.651
2	375.685	12	375.640
3	376.583	13	375.499
4	375.848	14	372.475
5	374.302	15	374.228
6	370.217	16	377.162
7	374.122	17	374.400
8	372.565	18	374.891
9	374.794	19	372.365
10	376.176	20	376.243

An estimate of the total amount of ^{235}U is required and must be within 0.5% of the unknown true value with probability 0.99. Determine the number of racks (clusters) that must be sampled using simple one-stage cluster sampling to meet the precision requirements.

The number m of racks (clusters) to be sampled from the $M = 20$ must be determined for the precision requirements $\epsilon = 0.005$ (0.5%) and $\gamma = 0.01$ ($z_{0.995} = 2.58$). The required sample size is given by Equation 8.14, where an estimate of CV_{clu}^2 is obtained from the information given in Table 8.2; that is,

$$\begin{aligned} (\hat{\text{CV}}_{\text{clu}})^2 &= \frac{(1/20)[(376.851 - \bar{X}_{\text{clu}})^2 + \dots + (376.243 - \bar{X}_{\text{clu}})^2]}{[(1/20)(376.851 + \dots + 376.243)]^2} \\ &= \frac{2.925}{(374.735)^2} = 0.00002083 \end{aligned}$$

Then the required number of racks to be sampled, calculated from Equation 8.14, is

$$m = \frac{(2.58)^2 (20) (0.00002083)}{(2.58)^2 (0.00002083) + (19)(0.005)^2} = 4.52$$

Thus, 5 of the 20 racks should be chosen by simple random sampling without replacement, and every container on each of the sampled racks must be measured.

Assume that the five racks selected have total (the sum of the measured values for all 30 containers on each rack) ^{235}U contents of 374.148, 376.492, 375.305, 373.198, and 374.315 kg. Then the desired estimate of the population total is obtained by applying Equation 8.12, which gives

$$\begin{aligned} x' &= N \bar{x} = N \left(\frac{M \sum_{i=1}^m x_i'}{m N} \right) \\ &= 600 \left[\frac{20(374.148 + 375.492 + 375.305 + 373.198 + 374.315)}{5(600)} \right] \\ &= 7489.832 \text{ kg} \end{aligned}$$

The standard error estimate is computed from Equation 8.13, which gives

$$\begin{aligned} \hat{SE}(x') &= \frac{M}{\sqrt{m}} \left(\frac{M-m}{M-1} \right)^{1/2} \left[\frac{\sum_{i=1}^m (x_i' - \bar{x}_{clu})^2}{m-1} \right]^{1/2} \left(\frac{M-1}{M} \right)^{1/2} \\ &= \frac{20}{\sqrt{5}} \left(\frac{20-5}{19} \right)^{1/2} \left(\frac{6.286}{4} \right)^{1/2} \left(\frac{19}{20} \right)^{1/2} = 9.710 \end{aligned}$$

An approximate $100(1 - \alpha)\%$ confidence interval for X' could be computed by substituting x' and $\hat{SE}(x')$ into Equation 8.44 (see Section 8.3.5).

8.3.2.2 Simple Two-Stage Cluster Sampling

Simple two-stage cluster sampling is used to select m clusters from the population of M clusters by simple random sampling and then to select n_i of the N_i containers within each cluster also by simple random sampling. This type of plan is not as economical as a simple one-stage cluster sampling plan because time is required to locate specific containers within each cluster. However, it does provide the flexibility of examining, at least partially, more clusters when the total number of containers in the sample is fixed. Also, when the containers within clusters are homogeneous, a more precise estimate may be obtained (perhaps more economically) by taking a sample of the containers within each of a larger number of clus-

ters than by measuring all containers within each of a smaller number of clusters.

For simple two-stage cluster sampling, the estimators for a population total and mean and their estimated standard errors are as follows:

Population Total, X'

$$x' = \frac{M}{m} \left(\sum_{i=1}^m \frac{N_i x'_i}{n_i} \right) \quad (8.15)$$

$$\begin{aligned} \hat{S}E(x') &= \frac{M(M-m)}{m(m-1)} \left[\sum_{i=1}^m (x'_i - \bar{x}_{clu})^2 \right. \\ &\quad \left. + \frac{M}{m} \sum_{i=1}^m \frac{N_i(N_i - n_i)}{n_i(n_i - 1)} \sum_{j=1}^{n_i} \left(x_{ij} - \frac{x'_i}{n_i} \right)^2 \right]^{1/2} \end{aligned} \quad (8.16)$$

Population Mean, \bar{X}

$$\bar{x} = \frac{x'}{N} \quad (8.17)$$

$$\hat{S}E(\bar{x}) = \frac{1}{N} \hat{S}E(x') \quad (8.18)$$

where

M = number of clusters in the population

m = number of clusters in the sample

N = number of containers in the population

N_i = number of containers in the i^{th} cluster

n_i = number of containers sampled from the i^{th} cluster

n = total number of containers in the sample

$x'_i = \sum_{j=1}^{n_i} x_{ij}$ = total value of all the containers sampled from the i^{th} sample cluster

x_{ij} = value of the variable of interest for the j^{th} container in the i^{th} cluster

$\bar{x}_{clu} = \sum_{i=1}^m x'_i / m$ = sample average of the cluster totals

Sample size formulas to achieve prespecified precision levels ϵ and γ with simple two-stage cluster sampling require choosing the number m of

clusters to sample in the first stage and the number n_i of containers to sample (in the second stage) from the i^{th} cluster chosen in the first stage. The formulas require more detailed information than is presented in this section, so they are not given here. Refer to Levy and Lemeshow (1980), Cochran (1977), or Yamane (1967). Beetle (1978) provides additional discussion of two-stage cluster sampling in nuclear material control applications.

8.3.3 Probability Proportional to Size (PPS) Sampling Plans

With probability proportional to size (PPS) sampling plans, the probability of selecting a particular container or cluster of containers for inclusion in the sample is based on some measure of size. The "size" of a container or cluster might be its SNM content or its strategic value. The choice of how size is measured depends on the specific application.

The development of a sampling plan with PPS is as follows. The first requirement is a listing of all containers or clusters in the population. In addition to information concerning the identity and location of all containers or clusters, the listing should include the size and cumulative size for each container or cluster in the population (see Example 8.4). The sizes should be expressed in appropriate size measurement units so that the size of each container or cluster is an integer. Denote the total number of size units in the population by N^* , an integer. Each container is associated with one or more size units, and each size unit is given an identification number between 1 and N^* . That is, a container or cluster of size y is associated with a particular set of y consecutive integers. Thus, each container or cluster has an identification set of integers.

Sampling begins by randomly selecting a number between 1 and N^* . Then the container† whose identification set contains the chosen random number is included in the sample. The same procedure used to select the first container could be used over and over, with the possibility that one or more containers might be chosen more than once. This is referred to as sampling *with* replacement. Sampling *without* replacement does not allow a container to be chosen more than once.

Sampling with replacement is easy to perform and provides a relatively simple theoretical background for the development of the desired estimating formulas. Allowing a container to be sampled more than once‡ can reduce the number of different containers in the sample. This is desirable because sampling costs are reduced, but it is undesirable because for a

†To simplify the discussion from this point on, the development will be presented for a population of containers. The development holds equally well for a population of clusters.

‡When a specific container is selected more than once, it is not necessary to remeasure it. Its one measurement result is given increased influence in the final population estimate due to its multiple selection.

fixed sample size the variance of the estimator of the population mean or total is larger than the corresponding variance for sampling without replacement.

8.3.3.1 PPS Sampling with Replacement

The estimators for a population total and mean and their estimated standard errors under PPS sampling with replacement are as follows:

Population Total, X'

$$x' = \sum_{i=1}^n \frac{x_i}{n\delta_i} \quad (8.19)$$

$$\hat{SE}(x') = \left[\frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{x_i}{\delta_i} - x' \right)^2 \right]^{1/2} \quad (8.20)$$

Population Mean, \bar{X}

$$\bar{x} = \frac{x'}{N} \quad (8.21)$$

$$\hat{SE}(\bar{x}) = \frac{1}{N} \hat{SE}(x') \quad (8.22)$$

where

x_i = value of the variable of interest for the i^{th} sample container

N = number of containers in the population

n = number of containers in the sample

$\delta_i = \frac{y_i}{N^*}$ = probability of selecting the i^{th} container to be in the sample

y_i = "size" of the i^{th} container (in size measurement units)

$N^* = \sum_{i=1}^N y_i$ = total size of the population (in size measurement units)

An approximate required sample size (for estimating a population mean or total) to achieve precision requirements ϵ and γ for PPS sampling with replacement is given by

$$n = \frac{(z_{1-\gamma/2})^2 \sum_{i=1}^N \delta_i \left(\frac{x_i}{\delta_i} - X' \right)^2}{(X')^2 \epsilon^2} \quad (8.23)$$

where

$$\epsilon = \text{maximum relative difference limit on } \frac{|x' - X'|}{X'} \text{ or } \frac{|\bar{x} - \bar{X}'|}{\bar{X}'}$$

$$\gamma = \text{probability that the absolute relative difference between the sample estimate and population value is greater than } \epsilon$$

In applications of Equation 8.23, the user is cautioned that the required sample size n can be driven to infinity by allowing ϵ to approach zero. Thus, it is possible for Equation 8.23 to yield a required sample size n that exceeds the population size N . This demonstrates an undesirable property of sampling with replacement; that is, in order to meet specified precision requirements, a sample size much larger than the population size may be required to ensure that adequate information from the population is included in the sample.

In nuclear material management applications, it would be ludicrous to take a sample that is larger than the population size N , especially when there is no guarantee that all N containers would be included in the sample. Thus, if Equation 8.23 yields a value of $n \geq N$ (or even if $n < N$, but n is moderately large relative to N), it is an indication that sampling without replacement would be a better strategy. In such cases, the sampling technique described in Section 8.3.3.2 should be considered.

Equation 8.23 is not immediately useful because it involves X' and x_i for $i = 1, 2, \dots, N$, that is, the unknown population total and its components. However, reasonable estimates from prior knowledge (such as from a recent census) may be substituted into Equation 8.23 to get an estimate of n . Also, it can be shown that the true population standard error of the estimator x' is given by

$$SE(x') = \left[\frac{1}{n} \sum_{i=1}^N \delta_i \left(\frac{x_i}{\delta_i} - X' \right)^2 \right]^{1/2}$$

so that Equation 8.23 can be written in the form

$$n = \frac{(z_{1-\gamma/2})^2(n)}{\epsilon^2} \left[\frac{SE(x')}{X'} \right]^2$$

Therefore, all that is needed to apply Equation 8.23 is an estimate of X' , say x'' , and an estimate of $SE(x'')$, say $\hat{SE}(x'')$, obtained from a previous sampling of the population where the sample size was n' . Then by substituting x'' , $\hat{SE}(x'')$, and n' into the right hand side of Equation 8.23, the approximate required sample size is given by

$$n \cong \frac{(z_{1-\gamma/2})^2(n')}{\epsilon^2} \left[\frac{\hat{SE}(x'')}{x''} \right]^2$$

With an appropriate change of notation, Equations 8.19 through 8.23 are applicable to single-stage PPS cluster sampling with replacement. The subscript i refers to the i^{th} cluster, while n and N are replaced with m and M . The estimators for a population total and mean and their estimated standard errors are as follows:

Population Total, X'

$$x' = \sum_{i=1}^m \frac{x_i'}{m \delta_i} \quad (8.24)$$

$$\hat{SE}(x') = \left[\frac{1}{m(m-1)} \sum_{i=1}^m \left(\frac{x_i'}{\delta_i} - x' \right)^2 \right]^{1/2} \quad (8.25)$$

Population Mean, \bar{X}

$$\bar{x} = \frac{x'}{M} \quad (8.26)$$

$$\hat{SE}(\bar{x}) = \frac{1}{M} \hat{SE}(x') \quad (8.27)$$

where

M = number of clusters in the population

m = number of clusters in the sample

$x_i' = \sum_{j=1}^{N_i} x_{ij} =$ total value of all containers in the i^{th} cluster

$N_i =$ number of containers in cluster i

$x_{ij} =$ value of the variable of interest for the j^{th} container in the i^{th} cluster

$\delta_i = y_i/M^* =$ probability of selecting the i^{th} cluster to be in the sample

$y_i =$ "size" of the i^{th} cluster (in size measurement units)

$M^* = \sum_{i=1}^M y_i =$ total size of the population of clusters (in size measurement units)

An approximate required sample size analogous to Equation 8.23 is given by

$$m = \frac{(z_{1-\gamma/2})^2 \sum_{i=1}^M \delta_i \left(\frac{x_i'}{\delta_i} - X' \right)^2}{(X')^2 \epsilon^2} \tag{8.28}$$

Equation 8.28 gives the number of clusters that must be sampled to meet the specified precision requirements. The caution and the discussion following Equation 8.23 are pertinent to applications of Equation 8.28.

Example 8.4 Cobb, Sapir, and Kern (1978) present a reference inventory for a large plutonium critical facility, where fuel plates are stored in one- or two-row drawers in the reactor cell, or in canisters in a vault. A summary of this inventory is displayed in Table 8.3. In this example, the drawers and canisters are considered as clusters of fuel plates.

TABLE 8.3
Reference Facility Inventory

Drawers	Pu, kg	Plates
Reactor: 1008 (1-row)	569	3024
576 (2-row)	651	3456
1584	1220	6480
Canisters	Pu, kg	Plates
Vault: 636 (full)	1772	8684
5 (partial)	8	45
641	1780	9729
Totals: M = 2225	X' = 3000	N = M* = 16,209

By sampling clusters of fuel plates with a PPS cluster sampling plan, estimate the total Pu content of the inventory to within 1% with probability 0.95. Use the number of fuel plates N_i in each cluster as the measure of cluster size. Note that δ_i , the probability of selecting the i^{th} cluster, is

given by $\delta_i = N_i/N$ (because $y_i = N_i$ and $M^* = N$, where N_i is the number of fuel plates in the i^{th} cluster and N is the total number of fuel plates in the inventory). Substituting N_i/N for δ_i results in the following formulas for this PPS single-stage cluster sampling problem:

$$x' = \frac{N}{m} \sum_{i=1}^m \frac{x_i'}{N_i}$$

$$\hat{SE}(x') = \left[\frac{N^2}{m(m-1)} \sum_{i=1}^m \left(\frac{x_i'}{N_i} - \frac{x'}{N} \right)^2 \right]^{1/2}$$

$$m = \frac{(z_{1-\gamma/2})^2 N \sum_{i=1}^M N_i \left(\frac{x_i'}{N_i} - \frac{X'}{N} \right)^2}{(X')^2 \epsilon^2}$$

where

$x_i' = \sum_{j=1}^{N_i} x_{ij}$ = total measured amount of Pu in the i^{th} cluster

x_{ij} = the measured amount of Pu in the j^{th} container from the i^{th} cluster

Because the individual cluster parameters x_i' and N_i are not given by Cobb, Sapir, and Kern (1978), average values computed from the summary information in Table 8.3 must be used. These are summarized below.

Number of clusters	Average x_i'	Average N_i
1008	$\frac{569}{1008} = 0.5645$	$\frac{3024}{1008} = 3$
576	$\frac{651}{576} = 1.1302$	$\frac{3456}{576} = 6$
636	$\frac{1772}{636} = 2.786$	$\frac{8684}{636} = 13.654$
5	$\frac{8}{5} = 1.600$	$\frac{45}{5} = 9$

The expression

$$\frac{N}{(X')^2} \sum_{i=1}^M N_i \left(\frac{x_i'}{N_i} - \frac{X'}{N} \right)^2$$

from the required sample size formula is then evaluated by substituting X' , N , and the average values of x_i' and N_i , which gives

$$\begin{aligned} & \frac{16,209}{(3000)^2} \left[(1008)(3) \left(\frac{0.5645}{3} - \frac{3000}{16,209} \right)^2 \right. \\ & + (576)(6) \left(\frac{1.1302}{6} - \frac{3000}{16,209} \right)^2 \\ & + (636)(13.654) \left(\frac{2.786}{13.654} - \frac{3000}{16,209} \right)^2 \\ & \left. + (5)(9) \left(\frac{1.600}{9} - \frac{3000}{16,209} \right)^2 \right] \\ & = \frac{16,209}{(3000)^2} (3.1903) = 0.00575 \end{aligned}$$

By substituting this value together with $\gamma = 0.05$ and $\epsilon = 0.01$ into the sample size formula, the approximate required sample size (number of clusters) is

$$m = \frac{(z_{1-\gamma/2})^2 (0.00575)}{\epsilon^2} = \frac{(1.96)^2 (0.00575)}{(0.01)^2} = 220.9 \cong 221$$

The selection of the 221 clusters (drawers or canisters) to be sampled proceeds as follows: 221 random integers between 1 and $M^* = 16,209$ are to be randomly selected with replacement. Using a cumulative size listing (number of fuel plates) such as that displayed in Table 8.4, the clusters associated with the 221 random numbers can be determined. For example, if 11 is one of the 221 random numbers chosen, cluster number 4 is in the sample because it contains the eleventh fuel plate. There may be fewer than 221 clusters in the final sample due to the possibility of some clusters being selected more than once. The clusters in the sample would then be measured and the results used to compute the estimate of

TABLE 8.4
Example Cumulative Listing

Cluster number	Cumulative number of plates
1	3
2	6
3	9
4	12
.	.
.	.
.	.
2,225	16,209

the total Pu content for the facility and its estimated standard error. A confidence interval for the facility Pu content could be calculated by using the formulas presented in Section 8.3.5.

8.3.3.2 PPS Sampling Without Replacement

The formulas for PPS sampling without replacement are somewhat more complicated than those for sampling with replacement, because each time a container is chosen to be in the sample, the probability of selecting each remaining container changes. Several simplified methods have been proposed by Cochran (1977). The Rao, Hartley, Cochran (RHC) method has several favorable features and is presented here.

The RHC method is carried out in two steps:

1. Split the population at random into n groups, of sizes N_1, N_2, \dots, N_n , where

$$\sum_{g=1}^n N_g = N$$

Values of the N_g are chosen to minimize the standard error of the estimator of the population total or mean (this is discussed further below).

2. Draw one container from each group with probability proportional to size sampling (within the group). The selection of a container within each group is performed independently of the other groups.

These two steps lead to a simple computational procedure for the estimates, because the probability of selecting any container remains fixed throughout the sample selection. Other sampling methods that skip Step 1 may be easier to carry out, but they lead to more complicated estimation formulas.

For the RHC method, the estimators for a population total and mean and their estimated standard errors are as follows:

Population Total, X'

$$x' = \sum_{g=1}^n \frac{x_g P_g}{\delta_g} \quad (8.29)$$

$$\hat{S}E(x') = \left\{ \frac{\sum_{g=1}^n N_g^2 - N}{N^2 - \sum_{g=1}^n N_g^2} \sum_{g=1}^n P_g \left(\frac{x_g}{\delta_g} - x' \right)^2 \right\}^{1/2} \quad (8.30)$$

Population Mean, \bar{X}

$$\bar{x} = \frac{x'}{N} \quad (8.31)$$

$$\hat{S}E(\bar{x}) = \frac{1}{N} \hat{S}E(x') \quad (8.32)$$

where

x_g = value of the variable of interest for the container sampled from the g^{th} group

n = number of containers in the sample

N = number of containers in the population

N_g = number of containers in the g^{th} group

$\delta_{gj} = \frac{y_{gj}}{N^*}$ = probability of selection (over the population) of the j^{th} container from the g^{th} group

y_{gj} = size of the j^{th} container in the g^{th} group

$N^* = \sum_{g=1}^n \sum_{j=1}^{N_g} y_{gj}$ = total size of the population

$P_g = \sum_{j=1}^{N_g} \delta_{gj}$ = the sum of the probabilities of selection for all containers in the g^{th} group

δ_g = the δ_{gj} value for the container selected from the g^{th} group

The standard errors given by Equations 8.30 and 8.32 are minimized by choosing the N_g such that

$$\begin{aligned}
 N_1 = \dots = N_n = \frac{N}{n} = R & \quad \text{if } R \text{ is an integer} \\
 N_1 = \dots = N_k = [R] + 1 & \left. \begin{array}{l} \text{if } R \text{ is not an integer, where} \\ N = n[R] + k, 0 < k < n, \text{ and} \\ [R] \text{ is the greatest integer in } R \end{array} \right\} \quad (8.33) \\
 \text{and} \\
 N_{k+1} = \dots = N_n = [R] &
 \end{aligned}$$

By implementing the RHC method, the sample is not selected with PPS over the whole population directly, but within the randomly formed groups. The method always provides an estimator for the population total or mean with smaller standard error than the estimators that result from sampling with replacement. Also, the standard error of the estimator decreases as the correlation between the measure of size y_{gi} and the true value of x_{gi} for a given container increases; in fact, the standard error is zero if the true value of x_{gi} and the measure of size y_{gi} are perfectly correlated. This is helpful because the previous recorded values of containers are often used as measures of size, so the measure of size and the current value are often approximately equal or at least highly correlated.

An approximate sample size required to achieve prespecified precision requirements ϵ and γ for the RHC method is given by

$$n = \frac{\frac{(z_{1-\gamma/2})^2 N}{(N-1)} \sum_{i=1}^N \delta_i \left(\frac{x_i}{\delta_i} - X' \right)^2}{(X')^2 \epsilon^2 + \frac{(z_{1-\gamma/2})^2}{(N-1)} \sum_{i=1}^N \delta_i \left(\frac{x_i}{\delta_i} - X' \right)^2} \quad (8.34)$$

where

$i = 1, 2, \dots, N$ is the index of the i^{th} container in the population without regard to groupings

x_i = value of the variable of interest for the i^{th} container in the population

δ_i = the δ_{gi} value, defined above, for the i^{th} container in the population

and other quantities are as previously defined. This formula is based on the optimal choice of the N_i as presented in Equation 8.33. This formula cannot be used unless X' and the quantity

$$J = \sum_{i=1}^N \delta_i \left(\frac{x_i}{\delta_i} - X' \right)^2 \quad (8.35)$$

are known or can be estimated. In practice, X' and J can often be replaced by reasonable estimates to provide an approximate required sample size.

It is interesting to compare Equation 8.34 with Equation 8.23. Note that if ϵ is set equal to zero, Equation 8.34 reduces to $n = N$, but Equation 8.23 yields $n = \infty$. This is because sampling *without* replacement guarantees that a sample of size $n = N$ will include the entire population; whereas sampling *with* replacement requires an infinite sample size to guarantee that the sample includes the entire population. In general, sampling without replacement requires a smaller sample size than sampling with replacement to achieve the same precision.

For cluster sampling, a slight modification of the notation is necessary. The estimator for the population total and its estimated standard error, given by Equations 8.29 and 8.30, are modified to

$$x' = \sum_{g=1}^m \frac{x'_g P_g}{\delta_g}$$

and

$$\hat{S}E(x') = \left[\frac{\sum_{g=1}^m M_g^2 - M}{M^2 - \sum_{g=1}^m M_g^2} \sum_{g=1}^m P_g \left(\frac{x'_g}{\delta_g} - x' \right)^2 \right]^{1/2}$$

where

x'_g = the total for the cluster sampled from the g^{th} group

m = number of clusters in the sample

M = number of clusters in the population

M_g = number of clusters in the g^{th} group

$\delta_{gj} = \frac{y_{gj}}{M^*}$ = probability of selection (over the population) of the j^{th} cluster from the g^{th} group

y_{gj} = "size" of j^{th} cluster in the g^{th} group

$M^* = \sum_{g=1}^m \sum_{j=1}^{M_g} y_{gj}$ = total size of the population

$P_g = \sum_{j=1}^{M_g} \delta_{gj}$ = the sum of the probabilities of selection for all clusters in the g^{th} group

δ_g = the δ_{gj} value for the cluster selected from the g^{th} group

The approximate required sample size formula, given by Equation 8.34, is modified to the form

$$m = \frac{\frac{(z_{1-\gamma/2})^2 M}{(M-1)} \sum_{i=1}^M \delta_i \left(\frac{x_i'}{\delta_i} - X' \right)^2}{(X')^2 \epsilon^2 + \frac{(z_{1-\gamma/2})^2}{(M-1)} \sum_{i=1}^M \delta_i \left(\frac{x_i'}{\delta_i} - X' \right)^2}$$

where

m = approximate number of clusters that must be in the sample to meet the precision requirements ϵ and γ

$i = 1, 2, \dots, M$ is the index of the i^{th} cluster in the population without regard to groupings

x_i' = total for the i^{th} cluster in the population

δ_i = the probability of selection of the i^{th} cluster from the population (see definition of δ_{gj} above)

The quantity J , given by Equation 8.35, is modified to

$$J = \sum_{i=1}^M \delta_i \left(\frac{x_i'}{\delta_i} - X' \right)^2$$

Example 8.5 Consider the large plutonium critical facility described in Example 8.4. Develop a PPS single-stage cluster sampling plan without replacement (using the RHC method) for estimating the total Pu content of the inventory. The precision requirements are $\gamma = 0.05$ and $\epsilon = 0.01$ as in Example 8.4.

The required sample size is calculated from Equation 8.34 with the appropriate cluster notation substituted. When the cluster notation is substituted, the quantity J (given by Equation 8.35) is equal to the summation in Equation 8.28. Thus, from Example 8.4, J is approximated by

51,711.6, and the total number of clusters is $M = 2225$. Then for $\gamma = 0.05$, $z_{0.975} = 1.96$, and the approximate required number of clusters is

$$m = \frac{(1.96)^2 (2225) (51,711.6)}{(2224)(3000)^2(0.01)^2 + (1.96)^2 (51,711.6)} = 200.9$$

Thus, 201 clusters are to be sampled. Note that this required sample size for PPS sampling without replacement is less than that required for sampling with replacement in Example 8.4, where $m = 221$. This is expected, because more clusters are necessary to meet the precision requirements when sampling with replacement to compensate for the potential loss of information due to clusters that may be sampled more than once.

The next step in the RHC procedure is to randomly split the population of $M = 2225$ clusters into $m = 201$ groups by implementing the procedure described in connection with Equation 8.33 as follows. First, note that

$$R = \frac{M}{m} = \frac{2225}{201} = 11.1$$

$$[R] = 11$$

$$m[R] = 201(11) = 2211$$

$$k = M - m[R] = 2225 - 2211 = 14$$

Then the population of 2225 clusters will be randomly split into 201 groups with the following sizes:

$$M_1 = \dots = M_{14} = 12$$

and

$$M_{15} = \dots = M_{201} = 11$$

This can be accomplished by randomizing the order of the integers from 1 to 2225 and letting the first 12 be Group 1, the second 12 be Group 2, and so on. Assume that this has been done and that Group 1 consists of the 12 clusters identified in Table 8.5.

One cluster is selected from this group by randomly choosing a number between 1 and 92. Then by using the cumulative size and cluster listing in Table 8.5, the cluster associated with the number chosen is included in the sample. For example, if the random number is 30, then cluster 1463 is included in the sample. This procedure is repeated for each of the 201

TABLE 8.5

**The First Group of Clusters for the RHC Method
of PPS Sampling Without Replacement**

Cluster identification	Size (No. plates)	Cumulative size (No. plates)	(δ)
1	15	3	3/16,209
2	173	3	6/16,209
3	462	3	9/16,209
4	628	3	12/16,209
5	937	3	15/16,209
6	1150	6	21/16,209
7	1280	6	27/16,209
8	1463	6	33/16,209
9	1754	15	48/16,209
10	1864	14	62/16,209
11	2148	14	76/16,209
12	2157	16	92/16,209

$P_1 = 92/16,209$

groups. An estimate of the population total and its estimated standard error can then be calculated by using Equations 8.29 and 8.30, with the modifications for cluster sampling.

8.3.4 Stratified Sampling Plans

It is often possible to stratify (partition into groups or classes) a population of containers on the basis of some characteristic, such as the type or amount of SNM in a container, or by measurement method (or variability). If a heterogeneous population can be divided into several subpopulations (strata) that are internally homogeneous with respect to the variable of interest, the precision of the estimate of a population mean or total will be improved compared to simple random sampling. However, the actual improvement in precision depends upon whether the variable of interest is correlated with the characteristic(s) used to stratify the population. If the correlation is high, then stratification can successfully improve the precision (reduce the standard error) of the estimate of the population mean or total. If the degree of homogeneity within each stratum is high, a precise estimate of the stratum mean or total can be obtained from a small sample within the stratum. Other properties of stratified sampling are listed in Table 8.1.

It may not always be possible to stratify a population of containers according to type or amount of SNM. If, for example, types of SNM and amounts in containers in a population vary widely, then the variance of a population mean or total estimate based on nonhomogeneous strata could be quite large, in fact, larger than the variance of estimates based on other sampling techniques. In such a situation, some of the previously discussed sampling techniques should be considered rather than the stratified sampling techniques discussed below.

8.3.4.1 How to Stratify a Population

A goal of stratification is to define strata that are internally homogeneous with respect to the variable of interest. This is accomplished by identifying one or more characteristics that are highly correlated with the variable of interest, and then using these characteristics as criteria for defining the strata. In nuclear material accounting applications, characteristics such as the type, amount, or measurement variability of SNM in a container might be used in defining strata. Often, measurement variability is highly correlated with the physical location of the container. This is because containers in the same location are usually measured by the same methods. Gains in precision of estimates resulting from stratification will be largest when the strata are considerably different from each other with respect to the variable of interest. Also, the number of strata should be kept small relative to the population size, because the total sample size required to meet precision goals will increase as the number of strata increases.† The task of stratifying a population depends on the characteristics of the population; a general procedure is outlined in Table 8.6.

The determination of group and subgroup boundaries in the procedure outlined in Table 8.6 has not been extensively discussed in the statistical and nuclear material safeguards literature. If only one stratification variable, such as the "size" of a container, is to be considered, Cochran (1977) suggests defining group boundaries by first generating a frequency distribution‡ of the population values of the size variable, and then calculating the square root of the frequency (\sqrt{f}) and the cumulative \sqrt{f} for each class interval. Group boundaries are then selected so that they yield equal intervals on the cumulative \sqrt{f} scale. For the case where two stratification variables are considered, such as size and type of SNM in a container, a similar procedure based on \sqrt{f} for each stratification variable is presented by Thomsen (1977).

†This criterion is presented from the practical standpoint that the costs of stratifying and analyzing the resulting sample data increase as the number of strata increase. Cochran (1977) and Hansen, Hurwitz, and Madow (1953) show that increasing the number of strata can decrease the standard error of the estimator, but that often a point of diminishing returns is reached very early (i.e., with a small number of strata), especially when cost is a major consideration.

‡See Chapters 1 and 2.

TABLE 8.6

Procedure for Stratification of a Population of SNM Containers

1. Obtain a complete list of all containers in the population. For each container, the list should include:
 - a. the type of SNM contained
 - b. the approximate amount (isotope or element) of SNM
 - c. the enrichment of the SNM
 - d. location information
 - e. expected measurement error and whether it varies on an absolute or relative basis.
 2. Group the containers by some measure of "size." Size might be an approximate isotope amount, element amount, or enrichment value for the container. This step is facilitated by forming a histogram for the size variable and then choosing group boundaries so that relatively homogeneous groups are obtained.
 3. If there are different types of SNM (such as UF_6 cylinders, UO_2 pellets, UO_2 powder, etc.) in the population, divide the groups generated in Step 2 into subgroups by material type. If all containers in the population hold the same type of SNM, the subgroups at this step are just the groups generated in Step 2.
 4. If there are significant differences among measurement error variances for containers within any Step 3 subgroup, subdivide that subgroup into smaller groups that are homogeneous with respect to measurement error variance. This step can sometimes alternately be performed by subdividing subgroups according to location, because containers stored in the same location are often measured by the same instrument, thus yielding the same measurement error variance for all containers at that location.
 5. Following Step 4, the population has been stratified.[†] The characteristics or definition of each stratum should be recorded for future use.
-

[†]Note that Steps 2, 3, and 4 can be permuted. For example, one might prefer to stratify first by SNM type and measurement variability (Steps 3 and 4) and then stratify by size (Step 2) last.

Another method of stratifying populations using one or more stratification variables is to use cluster analysis, a statistical technique which forms groups of homogeneous items (see Hartigan, 1975). Green, Frank, and Robinson (1967), Golder and Yeomans (1973), and Jarque (1981) discuss cluster analysis for stratification in specific problems.

Little appears in the literature about the effects of stratification when it is based on two or more stratification variables. Thomsen (1977) surmises from his work that

When one chooses to construct many strata, the gain from using two stratifying variables instead of one seems to be nontrivial when the correlations between the study variable and each of the stratifying variables are of some size, and the correlation between the two stratifying variables is small. Under the same conditions the results

indicate that for a given number of strata, it is more efficient to use two stratifying variables and make a few strata along each variable, as compared with using only the best stratifying variable and make optimal stratification along this variable.

If the Table 8.6 procedure is considered in light of the above comments, the three stratification variables—size, type, and measurement variance—satisfy some of the conditions, but not others. In most nuclear material control applications, the size variable is highly correlated with the variable of interest (SNM content), and hence contributes to successful stratification. While the type of SNM and the measurement variance may not contribute as much as the size, they can aid in structuring the problem to meet the assumptions required for specific techniques.

Given that the population is stratified, stratified sampling plans for estimating a population mean or total are discussed below.

8.3.4.2 Stratified Random Sampling

Stratified random sampling is merely simple random sampling within each stratum of the population. The estimators for a population mean and total and their estimated standard errors are as follows:

Population Mean, \bar{X}

$$\bar{x} = \sum_{i=1}^L \frac{N_i}{N n_i} \sum_{j=1}^{n_i} x_{ij} \quad (8.36)$$

$$\hat{SE}(\bar{x}) = \left[\sum_{i=1}^L \left(\frac{N_i}{N} \right)^2 \left(\frac{s_i^2}{n_i} \right) \left(\frac{N_i - n_i}{N_i} \right) \right]^{1/2} \quad (8.37)$$

Population Total, X'

$$x' = N \bar{x} \quad (8.38)$$

$$\hat{SE}(x') = N \hat{SE}(\bar{x}) \quad (8.39)$$

where

x_{ij} = value of the variable of interest for the j^{th} container in the i^{th} stratum

L = number of strata

N = number of containers in the population

N_i = number of containers in the i^{th} stratum

n_i = number of containers sampled from the i^{th} stratum

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1} = \text{sample variance of sampled containers within the } i^{\text{th}} \text{ stratum}$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \text{sample mean for the } i^{\text{th}} \text{ stratum}$$

The sample size required to achieve prespecified γ and ϵ for stratified random sampling can be approximated by considering the equation

$$\epsilon^2 (X')^2 = (z_{1-\gamma/2})^2 \sum_{i=1}^L N_i^2 \frac{\sigma_i^2 (N_i - n_i)}{n_i (N_i - 1)} \quad (8.40)$$

where

ϵ = maximum relative difference limit on $\frac{|x' - X'|}{X'}$ or $\frac{|\bar{x} - \bar{X}|}{\bar{X}}$

γ = probability that the absolute relative difference between the sample estimate and population value is greater than ϵ

$$\sigma_i^2 = \frac{\sum_{j=1}^{N_i} (x_{ij} - \bar{X}_i)^2}{N_i} = \text{population variance within stratum } i$$

Equation 8.40 is not immediately useful for two reasons. First, X' (the population total) and σ_i^2 (the population variance within the i^{th} stratum) are unknown. Second, the total sample size $n = \sum n_i$ cannot be determined explicitly. Further, there are many possible sets of n_i that satisfy Equation 8.40, and any one of the nonunique sets of n_i might be "better" than the others. In general, the best set depends upon the criterion chosen to make the decision. One criterion is to allocate n among the strata to minimize the population variance of the estimates x' or \bar{x} . The allocation of the n elements among the L strata is given by

$$n_i = n \left(\frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right) \quad (8.41)$$

where

$$n = \frac{(z_{1-\gamma/2})^2 \left(\sum_{i=1}^L \frac{N_i^2 \sigma_i}{N_i - 1} \right) \left(\sum_{i=1}^L N_i \sigma_i \right)}{\epsilon^2 (X')^2 + (z_{1-\gamma/2})^2 \sum_{i=1}^L \frac{N_i^2 \sigma_i^2}{N_i - 1}} \quad (8.42)$$

To use Equations 8.41 and 8.42, the population parameters X' and σ_i must be replaced with reasonable estimates.

Equations 8.41 and 8.42 do not account for the cost of sampling and measuring a container within a stratum. Cochran (1977) gives a cost function and related formulas that can be applied to determine a sample size and optimum allocation among the strata to either minimize the variance subject to a fixed budget or minimize costs subject to specified precision requirements.

Example 8.6 Suppose that the large plutonium critical facility in Example 8.4, is considered to be a stratified population of containers as follows:†

Stratum	Number of containers, N_i	Pu, kg
1-Row reactor drawers	1008	569
2-Row reactor drawers	576	651
Full-vault canisters	636	1772
Partial-vault canisters	5	8
Total	2225	3000

Develop a stratified random sampling plan for estimating the population total Pu, with precision requirements $\gamma = 0.05$ and $\epsilon = 0.01$.

The total sample size and its allocation among strata are calculated from Equations 8.42 and 8.41. Assume that reasonable estimates for σ_i , $i = 1, 2, 3, 4$, in kg Pu are available from past data and are displayed in Table 8.7. Then the required total sample size n over all strata is computed from Equation 8.42, where

†In this example the drawers and canisters are not considered to be clusters as in Examples 8.4 and 8.5.

$$\sum_{i=1}^L \frac{N_i^2 \sigma_i}{N_i - 1} \cong \frac{(1008)^2(0.047)}{1007} + \frac{(576)^2(0.104)}{575} \\ + \frac{(636)^2(0.215)}{635} + \frac{(5)^2(0.352)}{4} = 246.59$$

$$\sum_{i=1}^L N_i \sigma_i \cong 1008(0.047) + 576(0.104) + 636(0.215) \\ + 5(0.352) = 245.78$$

$$\sum_{i=1}^L \frac{N_i^2 \sigma_i^2}{N_i - 1} \cong \frac{(1008)^2(0.047)^2}{1007} + \frac{(576)^2(0.104)^2}{575} \\ + \frac{(636)^2(0.215)^2}{635} + \frac{(5)^2(0.352)^2}{4} = 38.69$$

This gives

$$n = \frac{(1.96)^2(246.59)(245.78)}{(0.01)^2(3000)^2 + (1.96)^2(38.69)} = 222.03 \cong 222$$

TABLE 8.7
Estimates of σ_i from Past Data

i	$\hat{\sigma}_i$, kg Pu
1	0.047
2	0.104
3	0.215
4	0.352

This total sample size of 222 containers (drawers or canisters) is allocated to the strata by applying Equation 8.41 as follows:

$$n_1 = 222 \left[\frac{1008(0.047)}{245.78} \right] = 42.8 \cong 43$$

$$n_2 = 222 \left[\frac{576(0.104)}{245.78} \right] = 54.1 \cong 54$$

$$n_3 = 222 \left\lceil \frac{636(0.215)}{245.78} \right\rceil = 123.5 \cong 124$$

$$n_4 = 222 \left\lceil \frac{5(0.352)}{245.78} \right\rceil = 1.6 \cong 2$$

Note that because of rounding in computing the n_i , the total sample size is 223 rather than 222. A sample of size n_i is to be taken from within the i^{th} stratum by simple random sampling. Following selection and measurement of the samples, Equations 8.38 and 8.39 can be applied to compute an estimate of the population total and its estimated standard error.

8.3.4.3 Other Stratified Sampling Plans

After a population has been stratified, there is no restriction on the type of sampling techniques that can be used to obtain a sample from within each stratum. However, for a particular application, sampling techniques should be chosen that are best suited to the problem at hand. Examples of a few stratified sampling plans that may be suited to nuclear material control problems are as follows:

1. Probability proportional to size (PPS) sampling of containers within strata.
2. Single-stage cluster sampling within strata, with clusters chosen by either simple random sampling or by PPS sampling.
3. Two-stage cluster sampling within strata with the first-stage cluster sampled as in (2) and the second-stage samples of containers from within clusters chosen either by simple random sampling or by PPS sampling.

Many other sampling plans may also be suitable. In fact, it is not necessary to use the same sampling technique within all strata. For example, in a particular application, simple random sampling might be most appropriate for some strata, but PPS cluster sampling might be most appropriate for other strata.

The main difficulty with using a technique other than simple random sampling within all strata, or with mixing sampling techniques across strata, is that formulas are not always readily available for computing estimates of means, totals, and standard errors. Approximation of the sample size(s) needed to meet specified precision requirements may also be difficult. However, these difficulties should not preclude use of this approach. Familiarity with the basic sampling techniques is sufficient to formulate a sampling plan tailored to a specific application. A statistician who specializes in sampling theory can then derive the necessary formulas.

8.3.5 Confidence Intervals for Population Means and Totals

Approximate two-sided $100(1 - \alpha)\%$ confidence intervals for the population mean (\bar{X}) and total (X') for each of the sampling plans in Sections 8.3.1 through 8.3.4 are given by:

$$\bar{x} \pm z_{1-\alpha/2} \hat{SE}(\bar{x}) \quad (8.43)$$

and

$$x' \pm z_{1-\alpha/2} \hat{SE}(x') \quad (8.44)$$

where

- \bar{x} = sample estimate of the population mean
- x' = sample estimate of the population total
- $\hat{SE}(\bar{x})$ = estimated standard error of \bar{x}
- $\hat{SE}(x')$ = estimated standard error of x'

and where $z_{1-\alpha/2}$ is obtained from Table A3. The confidence intervals given by Equations 8.43 and 8.44 are approximate because the value $z_{1-\alpha/2}$ is obtained from the standard normal distribution and not from the actual distribution of x' or \bar{x} . Because x' and \bar{x} are computed by summing over observed values of a random variable, the normal distribution approximates the actual distributions of x' and \bar{x} , and the accuracy of the approximation improves as the sample size increases. For more detail, see the discussion of the Central Limit Theorem in Section 2.5.2.1.

REFERENCES

- Beetle, T. M., 1978, Two-Stage Sampling in Safeguards Work, *International Symposium on Nuclear Safeguards Technology*, Vol. II, IAEA-SM-231/103, International Atomic Energy Agency, Vienna.
- Cobb, D. C., J. L. Sapir, and E. A. Kern, 1978, Inventory Verification in Large Critical Facilities, *J. Inst. Nucl. Mater. Manage.*, (Proc. Issue), VII: 233-242.
- Cochran, W. G., 1977, *Sampling Techniques*, 3rd Ed., John Wiley & Sons, Inc., New York.
- Golder, P. A., and K. A. Yeomans, 1973, The Use of Cluster Analysis for Stratification, *Applied Statistics*, 22: 213-219.
- Green, P. E., R. E. Frank, and A. G. Robinson, 1967, Cluster Analysis in Test Market Selection, *Management Science*, 13(8): 387-400.
- Hansen, M. H., W. N. Hurwitz, and W. G. Madow, 1953, *Sample Survey Methods and Theory—Volume I Methods and Applications*, John Wiley & Sons, Inc., New York.
- Harkness, A. L., 1977, The Effort of Obtaining a Random Sample, *J. Inst. Nucl. Mater. Manage.*, VI(1): 34-36.

- Hartigan, J. A., 1975, *Clustering Algorithms*, John Wiley & Sons, Inc., New York.
- Jarque, C. M., 1981, A Solution to the Problem of Optimum Stratification in Multivariate Sampling, *Applied Statistics*, **30**: 163-169.
- Levy, P. S., and S. Lemeshow, 1980, *Sampling for Health Professionals*, Wadsworth, Inc., Belmont, California.
- Piepel, G. F., and R. J. Brouns, 1982, *Statistical Sampling Plans for Prior Measurement Verification and Determination of the SNM Content of Inventories*, NUREG/CR-2466 (PNL-4057), Pacific Northwest Laboratory, Richland, Washington.
- Thomsen, I., 1977, On the Effect of Stratification When Two Stratifying Variables Are Used, *J. Am. Stat. Assoc.*, **72**: 149-153.
- Yamane, T., 1967, *Elementary Sampling Theory*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

CHAPTER 9

Nonparametric Methods

9.0 INTRODUCTION

This chapter presents nonparametric statistical methods that are counterparts to some of the parametric methods presented in the previous chapters, such as t-tests, analysis of variance, and correlation. The basic distinction is that parametric procedures require specific assumptions about the probability distributions from which empirical observations arise; the most common distributional assumption is that the observations are from a normal distribution. Nonparametric methods require less restrictive distributional assumptions. When specific parametric assumptions are true, however, this relaxation of assumptions usually results in less powerful hypothesis tests than would result if the appropriate parametric methods were used. But when the parametric assumptions are not true, the parametric methods can yield misleading results, while the nonparametric methods can provide valid results. Another advantage of nonparametric methods is that the practitioner can determine the exact probabilities of the Type I and Type II errors for most nonparametric hypothesis testing methods. For many specific types of problems, Conover (1980) provides in-depth comparisons of alternative nonparametric methods with each other and with their parametric counterparts.

Nonparametric techniques are useful in situations where the probability distribution from which the observations were taken is either unknown or is known not to be normal. While many parametric techniques are fairly robust (not sensitive to incorrect assumptions), some are not. Thus, in some situations, nonparametric statistical methods are preferred. Even so, because of a general limited awareness, nonparametric methods are not used as extensively as they could be in nuclear material accounting (and many other) applications.

Most of the methods presented in this chapter are based on replacing observations by their ranks or by zero-one (yes-no or plus-minus) values. The procedures based on ranks or zero-one data are more robust than their parametric counterparts because interval or ratio data are transformed to ordinal† data; thus, outlying observations are not given undue weight.

†See Section 1.1 for a discussion of nominal, ordinal, interval, and ratio measurement scale data.

When interval or ratio data are transformed to ordinal data, however, some information is sacrificed; this is why nonparametric hypothesis testing methods are generally less powerful than their parametric counterparts when the parametric assumptions are valid.

The above discussion introduces nonparametric statistical methods as alternates to the parametric methods when distributional assumptions are not satisfied. This characterization is certainly appropriate for problems where the underlying probability distribution is some continuous distribution other than the normal distribution. For other situations, however, even the assumption of a continuous distribution may not be appropriate. As discussed in Section 1.1, data may be measured on a nominal or ordinal scale, as opposed to an interval or ratio scale. In some cases the parametric methods may not be applicable and nonparametric methods (such as those presented in this chapter) provide the only valid approach to making inferences from the data.

The reader is cautioned that nonparametric methods are not completely free of distributional assumptions. Most nonparametric procedures require that measured values be independent and/or identically distributed. In many safeguards applications, such as those involving inventory differences or measurements with particular types of multiplicative error structures, these assumptions are frequently inappropriate. As with parametric methods, the assumptions underlying nonparametric methods should be carefully checked before they are applied.

Sections 9.1 through 9.4 present nonparametric techniques for testing hypotheses about location[†] and/or dispersion (variance) of one or more populations, where the samples from the populations can be either related or independent. These techniques are the nonparametric analogs of the hypothesis testing and analysis of variance methods presented in Chapters 4 and 5. Section 9.5 discusses nonparametric correlation techniques. Section 9.6 covers methods for testing distributional assumptions, such as the assumption of normality.

9.1 SINGLE SAMPLE OR MATCHED PAIRS

A test for randomness of observations in a single sample is introduced in this section. Methods are also presented for testing hypotheses about location using data from a single sample or using matched pairs.

9.1.1 A Test of Randomness: The Runs Test

Most parametric and nonparametric techniques are based on the assumption of random samples. Inferences about populations are based on

[†]In this chapter, the term "location" generally refers to a measure of central tendency, such as the mean or median. Although this term is adopted from the "location parameter" terminology of probability distributions, not all location parameters are related to the concept of central tendency.

samples collected from the population; if the samples are not randomly selected from the populations under study, erroneous and misleading conclusions could be reached. Because the assumption of a random sample is an important one, it should be checked in most applications. The runs test is a procedure that can be used to check this assumption.

Consider a series of observations recorded in the order they were observed. Replace each observation by a "+" if it is greater than the sample median (Section 1.2.3) and by a "-" if it is less than or equal to the median.† A random series of observations should produce a randomly intermixed string of plus and minus signs such as

+ + + - - + - + + - - - +

The randomness of patterns such as

+ + + + + + + - - - - - - -

or

+ - + - + - + - + - + - +

would be highly suspect. The runs test for randomness uses the number of runs (sequences of the same sign) of plus or minus signs. If there are too many or too few runs compared to the number that would be expected from a random sample, the null hypothesis of randomness is rejected. Under the null hypothesis of randomness, the number of runs is a random variable that has a probability distribution, and the statistical test of the hypothesis is based on comparing the observed number of runs with quantiles of this distribution.

Let n be the total number of signs (observations), with n_1 plus signs and n_2 minus signs ($n_1 + n_2 = n$). For n_1 and $n_2 \leq 10$, Swed and Eisenhart (1943) calculated probabilities for the number of runs U . A table of these cumulative probabilities, compiled by Draper and Smith (1981), is given in Table A11 in Appendix A. Let the observed number of runs for the sample under investigation be U^* . Then, for a specified significance level α , the null hypothesis of randomness is rejected if

$$\Pr(U \leq U^*) \leq \frac{\alpha}{2}$$

†Testing for randomness is a special application of the runs test. Reference values other than the median are used for other applications.

or if

$$\Pr(U \geq U^*) = 1 - \Pr(U \leq U^* - 1) \leq \frac{\alpha}{2} \quad (9.1)$$

where $\Pr(U \leq u)$ is obtained from Table A11 for n_1 and $n_2 \leq 10$.

Example 9.1 To estimate an average percent by weight uranium value for a production lot of UO_2 powder, analyses are made on 14 samples drawn as the powder was produced. The percent uranium values (in order) are:

87.590, 87.588, 87.595, 87.603, 87.601, 87.582, 87.580,
87.597, 87.599, 87.609, 87.601, 87.610, 87.611, 87.615

Test the hypothesis that these observations are a random sample from the production lot. Use the runs test with the $\alpha = 0.05$ level of significance.

Because there are an even number of observations, the median is calculated as the average of the middle two (ordered) observations 87.599 and 87.601, yielding a median of 87.600. The observations are replaced by “+” or “-” depending on whether they are greater than 87.600 or less than or equal to 87.600, respectively. For example, the first three values are less than 87.600, and they are replaced with minus signs. The fourth and fifth values are greater than 87.600, and they are replaced with plus signs. This process yields the sequence

- - - + + - - - + + + + +

There are $n_1 = 7$ plus signs and $n_2 = 7$ minus signs, with $U^* = 4$ runs. Then, by using Equation 9.1 and Table 11,

$$\Pr(U \leq 4) = 0.025$$

and

$$\Pr(U \geq 4) = 1 - \Pr(U \leq 3) = 1 - 0.004 = 0.996$$

Because $\Pr(U \leq 4) = 0.025 \leq 0.05/2 = 0.025$, the null hypothesis of randomness is rejected at the $\alpha = 0.05$ level of significance. Fewer runs

were observed than would be expected if the null hypothesis of randomness were true. This implication of nonrandomness is important from an accountability viewpoint, because an average percent uranium factor for the lot has less meaning when the observations are not a random sample from the production lot.

When n_1 and $n_2 > 10$ for the runs test, exact critical values are not provided because a normal approximation to the actual distribution is satisfactory. In such cases, the statistic

$$Z^* = \frac{U^* - \mu + a}{\sigma} \quad (9.2)$$

has an approximate standard normal distribution where μ and σ are the actual mean and standard deviation of the discrete distribution of U under the null hypothesis of randomness. It can be shown that

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad (9.3)$$

and

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \quad (9.4)$$

Also,

$$a = \begin{cases} 0.5 & \text{when testing for "too few runs"} \\ -0.5 & \text{when testing for "too many runs"} \end{cases}$$

The 0.5 or -0.5 in Equation 9.2 is a **continuity correction** that helps to compensate for a continuous distribution being used to approximate a discrete distribution (e.g., Section 2.5.2.2).

A two-sided test of randomness with significance level α is performed by comparing the test statistics $Z^*_{\text{too many}}$ and $Z^*_{\text{too few}}$ against the standard normal values $z_{\alpha/2}$ and $z_{1-\alpha/2}$, obtained from Table A3. Note that $Z^*_{\text{too many}}$ is obtained by substituting $a = -0.5$ into Equation 9.2 and $Z^*_{\text{too few}}$ is obtained by using $a = 0.5$. The null hypothesis of randomness is rejected if either $Z^*_{\text{too many}} > z_{1-\alpha/2}$ or $Z^*_{\text{too few}} < z_{\alpha/2}$.

Example 9.2 For a problem similar to that in Example 9.1, analyses on 25 samples were performed with $n_1 = 12$, $n_2 = 13$, and $U^* = 17$. Test the hypothesis of randomness for the associated sample at the $\alpha = 0.10$ level of significance.

The normal approximation is used with the test statistic Z^* computed from Equation 9.2, where

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{2(12)(13)}{12 + 13} + 1 = 13.48$$

$$\begin{aligned}\sigma &= \left[\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \right]^{1/2} \\ &= \left[\frac{2(12)(13)[2(12)(13) - 12 - 13]}{(12 + 13)^2(12 + 13 - 1)} \right]^{1/2} \\ &= \sqrt{5.97} = 2.44\end{aligned}$$

Then

$$Z^*_{\text{too many}} = \frac{17 - 13.48 - 0.5}{2.44} = 1.24$$

and

$$Z^*_{\text{too few}} = \frac{17 - 13.48 + 0.5}{2.44} = 1.65$$

The values $z_{0.05} = -1.65$ and $z_{0.95} = 1.65$ are obtained from Table A3. Because $Z^*_{\text{too few}} = 1.65 > -1.65$, the "too few" test does not lead to rejection of the null hypothesis. This is logical, because there are more runs (17) than the expected number (13.48). Also, because $Z^*_{\text{too many}} = 1.24 < 1.65$, the null hypothesis of randomness is not rejected. Thus, it is appropriate to calculate an average percent uranium factor for the production lot based on these observations.

Another application of the runs test is to test for **serial correlation** (autocorrelation) in a sequence of observations. This application is discussed in Section 6.4.5.3 and by Draper and Smith (1981). Other tests for randomness besides the runs test are presented by Lehmann (1975) and Conover (1980).

9.1.2 Single Sample Tests of Location

This section presents two nonparametric methods for testing hypotheses about location based on a single sample: the sign test and the Wilcoxon signed ranks test. These tests can be applied to investigate the validity of the assumption that a sample comes from a population with a specified value m of the median;† that is, to test the hypothesis

$$H_0: \text{Median} = m$$

against the alternative

(9.5)

$$H_A: \text{Median} \neq m$$

These tests are the nonparametric analogs of the single sample t -test presented in Section 4.2.2.

9.1.2.1 The Sign Test

It is assumed that a random sample has been taken from a single population and that the characteristic of interest is measured on at least an ordinal scale (Section 1.1). Each observation is compared to the hypothesized value m of the median and is replaced with a “+” if the observation is larger than m , and with a “-” if it is smaller than m . Observations equal to the median are discarded and the sample size n' is correspondingly reduced.‡ Let n represent the reduced sample size.

The test statistic T^* is the total number of plus signs. A two-sided test is performed by determining whether this observed number of plus signs is too large or too small compared to the number that would be expected when the null hypothesis is true. When H_0 in Equation 9.5 is true, the random variable T defined as the “number of plus signs” has a binomial dis-

†The sign test and Wilcoxon signed ranks test can also be applied to test hypotheses about the mean if the underlying probability distribution is symmetric.

‡Lehmann (1975) refers to the sign test with observations equal to the median discarded as a conditional sign test and notes that it results in a more powerful test than the random tiebreaking rule often recommended for applications of the sign test.

tribution (Section 2.4.2) with $p = \frac{1}{2}$ and sample size n . When $n > 20$, the normal approximation to the binomial (Section 2.5.2.2) is used to approximate the distribution of T .

The sign test is performed as follows. For a specified value of α , the null hypothesis is rejected if

$$\Pr(T \leq T^*) \leq \frac{\alpha}{2}$$

or

$$\Pr(T \geq T^*) = 1 - \Pr(T \leq T^* - 1) \leq \frac{\alpha}{2} \quad (9.6)$$

For the one-sided alternative H_A : Median $> m$, reject H_0 if $\Pr(T \geq T^*) \leq \alpha$. For the one-sided alternative H_A : Median $< m$, reject H_0 if $\Pr(T \leq T^*) \leq \alpha$. If $n \leq 20$, the cumulative binomial distribution (Table A1) is used to find the above probabilities. If $n > 20$, the normal approximation to the binomial (Section 2.5.2.2) is used.

Example 9.3 Periodically, an NDA (gamma scan) instrument is checked for bias by measuring the grams ^{235}U in a solid waste barrel working standard (with an assigned value of 2.40 g). During a 6 month period, the standard is measured 12 times with the following results.

2.92, 2.35, 2.44, 2.21, 2.58, 2.65,
2.73, 2.04, 2.42, 2.62, 2.17, 2.39

Test the hypothesis that the population median is equal to 2.40 g. Use the sign test with the $\alpha = 0.10$ level of significance.

Each observation is replaced by a "+" or "-" depending upon whether it is greater than or less than 2.40. The following sequence results

+ - + - + + + - + + - -

The sample appears to be random, which can be verified with the runs test; thus, the sign test is applicable. The total number of plus signs is $T^* = 7$. Using Equation 9.6 and the cumulative binomial distribution (Table A1) with $n = 12$ and $p = \frac{1}{2}$, the values for the two-sided test are

$$P(T \leq 7) = 0.8062$$

and

$$P(T \geq 7) = 1 - P(T \leq 6) = 1 - 0.6128 = 0.3872$$

Because $0.8062 > 0.05$ and $0.3872 > 0.05$, the null hypothesis is not rejected at the $\alpha = 0.10$ level of significance. There is no significant indication that the NDA instrument is biased relative to the working standard.

9.1.2.2 The Wilcoxon Signed Ranks Test

The Wilcoxon signed ranks test can be used as a test of location for a single population, although it has other uses as well. The measure of location can be either the median or the mean, as discussed below. Let x_1, x_2, \dots, x_n be a random sample of n' observations of the random variable X , where the characteristic of interest is measured on at least an ordinal measurement scale and X has a symmetric distribution. It is of interest to test the hypothesis

$$H_0: \text{Median (or Mean)} = m$$

against the alternative

$$(9.7)$$

$$H_A: \text{Median (or Mean)} \neq m$$

Note that a hypothesis about the mean is equivalent to a hypothesis about the median when the distribution of X is symmetric.

The sample observations are ordered from smallest to largest, and the quantity $d_i = m - x_i$ is computed for each observation x_i , $i = 1, 2, \dots, n'$. Omit from further consideration all observations yielding $d_i = 0$ (i.e., where $x_i = m$). Denote the number of nonzero d_i values by n and the absolute value of d_i by $|d_i|$. Ranks from 1 to n are assigned to the $|d_i|$ values (assign the rank of 1 to the smallest $|d_i|$ and n to the largest). If two or more $|d_i|$ values are equal, each value is assigned a rank equal to the average of the ranks that would otherwise have been assigned. Then, let $R_i = 0$ if d_i is negative, and $R_i =$ the rank of $|d_i|$ if d_i is positive. The test statistic is given by

$$W^* = \sum_{i=1}^n R_i \quad (9.8)$$

A value of W^* that is either too small or too large indicates that the null hypothesis should be rejected. Specifically, H_0 is rejected at the α level of significance if W^* exceeds $w_{1-\alpha/2}$ or if W^* is less than $w_{\alpha/2}$, where w_p is obtained from Table A12. For the one-sided alternative H_A : Median (or mean) $< m$, reject H_0 if $W^* > w_{1-\alpha}$. For the one-sided alternative H_A : Median (or Mean) $> m$, reject H_0 if $W^* < w_{\alpha}$.

Use of the Wilcoxon signed ranks test to test a hypothesis about location using data from a single population is illustrated in the following example.

Example 9.4 Use the Wilcoxon signed ranks test and the data from Example 9.3 to test the hypothesis that the median is 2.40 g. Use the $\alpha = 0.10$ level of significance.

The ordered observations, the differences $d_i = 2.40 - x_i$, the ranks of the $|d_i|$, and the values of R_i are displayed in Table 9.1. The test statistic, computed from Equation 9.8, is

$$W^* = \sum_{i=1}^{12} R_i = 30$$

TABLE 9.1
Observations and Ranking Results

| x_i | $d_i = 2.40 - x_i$ | Rank
of $ d_i $ | R_i |
|-------|--------------------|--------------------|-------|
| 2.04 | 0.36 | 11 | 11 |
| 2.17 | 0.23 | 8 | 8 |
| 2.21 | 0.19 | 6 | 6 |
| 2.35 | 0.05 | 4 | 4 |
| 2.39 | 0.01 | 1 | 1 |
| 2.42 | -0.02 | 2 | 0 |
| 2.44 | -0.04 | 3 | 0 |
| 2.58 | -0.18 | 5 | 0 |
| 2.62 | -0.22 | 7 | 0 |
| 2.65 | -0.25 | 9 | 0 |
| 2.73 | -0.33 | 10 | 0 |
| 2.92 | -0.52 | 12 | 0 |

For $n = 12$ and $\alpha = 0.10$, the values $w_{0.05} = 18$ and $w_{0.95} = [12(13)2] - 18 = 60$ are obtained from Table A12. Because $18 < 30 < 60$, the null hypothesis is not rejected at the $\alpha = 0.10$ level of significance. There is insufficient evidence to conclude that the median differs from 2.40.

Two single sample nonparametric methods have been considered for testing hypotheses about location: the sign test and the Wilcoxon signed ranks test. These are counterparts of the single sample t-test discussed in Section 4.2.2. There is a hierarchy in the assumptions required for these three tests. The sign test assumes a random sample from the population. The Wilcoxon signed ranks test adds the assumption that the distribution is symmetric about the median (or mean). The t-test adds the assumption that the distribution is normal. The choice of which test to use should be based on how realistic the assumptions of each test are for the problem at hand. If the assumptions for a given test are valid, that test is more powerful for detecting deviations from H_0 than the tests lower in the hierarchy (i.e., tests requiring less restrictive assumptions).

9.1.3 Matched Pairs Tests for Comparing Locations

A common task in nuclear material control involves verifying prior information on a population of SNM containers. One such task is to investigate the possibility of a shipper-receiver difference; another involves an inspector verifying operator data. A common quality control situation involves a comparison of two analytical laboratories (or instruments, technicians, etc.) by having them each measure identical samples of several items. The appropriate statistical test for these situations is one that compares location parameters (means or medians) on the basis of paired observations.

The parametric method for testing the hypothesis of equal locations using paired observations is the paired t-test (Section 4.4). The paired t-test is conducted by taking the difference between the paired observations and then performing a single sample t-test by using the differences as data. The nonparametric matched pairs tests of location are conducted in a similar manner; that is, depending on the assumptions made, either the sign test or the Wilcoxon signed ranks test can be applied to the differences between the paired observations. Note that when the differences are used as data, they replace the d_i in the Wilcoxon test, and when the signs of the differences are used as data, then the sign test is applied. Application of the Wilcoxon signed ranks test with paired observations is illustrated in Example 9.5.

Example 9.5† When plutonium nitrate solution is being loaded into a recovery plant for further purification, samples are drawn from each container and analyzed for percent plutonium using one of two different analytical methods. One analytical method is a direct assay, and the other requires that a correction factor be applied. Periodically, the equivalence of the two methods is checked. For this purpose, samples from 20 containers were recently analyzed by both methods. Using the resulting data displayed in Table 9.2, test the hypothesis that the plutonium percentages are the same for both analytical methods. Use the $\alpha = 0.10$ level of significance.

TABLE 9.2
Plutonium Concentration Data (% of Pu)

| Container | Method 1 | Method 2 | Method 1-Method 2
(d_i) | Ranks of
$ d_i $ | R_i |
|-----------|----------|----------|--------------------------------|---------------------|---------|
| 1 | 13.11 | 13.00 | 0.11 | 11 | 11 |
| 2 | 15.14 | 14.90 | 0.24 | 16 | 16 |
| 3 | 13.22 | 13.01 | 0.21 | 15 | 15 |
| 4 | 13.67 | 13.65 | 0.02 | 3.5 | 3.5 |
| 5 | 10.48 | 10.61 | -0.13 | 12.5 | 0 |
| 6 | 15.37 | 15.11 | 0.26 | 19 | 19 |
| 7 | 12.37 | 12.40 | -0.03 | 6 | 0 |
| 8 | 12.50 | 12.63 | -0.13 | 12.5 | 0 |
| 9 | 11.46 | 11.71 | -0.25 | 17.5 | 0 |
| 10 | 14.28 | 14.21 | 0.07 | 9 | 9 |
| 11 | 13.26 | 13.01 | 0.25 | 17.5 | 7.5 |
| 12 | 11.00 | 11.06 | -0.06 | 8 | 0 |
| 13 | 12.74 | 12.75 | -0.01 | 1.5 | 0 |
| 14 | 13.69 | 13.69 | 0 | omitted | omitted |
| 15 | 10.43 | 10.40 | 0.03 | 6 | 6 |
| 16 | 11.38 | 11.30 | 0.08 | 10 | 10 |
| 17 | 12.26 | 12.27 | -0.01 | 1.5 | 0 |
| 18 | 12.89 | 12.70 | 0.19 | 14 | 14 |
| 19 | 13.33 | 13.30 | 0.03 | 6 | 6 |
| 20 | 11.88 | 11.90 | -0.02 | 3.5 | 0 |

The problem is to test the null hypothesis $H_0: m_1 = m_2$ against the two-sided alternative $H_A: m_1 \neq m_2$, where m_i is the unknown true median percentage Pu for the i^{th} analytic technique. The results are obviously paired, because a sample from each container was analyzed by both techniques.

†The data for this example are from Jaech (1973).

The percent Pu differences d_i are given in Table 9.2. Suppose that from past experience it is reasonable to assume that the d_i have a symmetric distribution but the distribution is not normal. If the null hypothesis is true, then the median of the d_i is zero. Thus, the Wilcoxon signed ranks test for a zero median is applied. In this case, note that the medians m_i in the hypothesis can be equivalently replaced by means μ_i . The ranks of the $|d_i|$ and the values of R_i are displayed in Table 9.2. The test statistic, calculated from Equation 9.8, is

$$W^* = \sum_{i=1}^{19} R_i = 117$$

For the two-sided test with $\alpha = 0.10$ and $n = 19$, the values $w_{0.05} = 54$ and $w_{0.95} = 190 - 54 = 136$ are obtained from Table A12. Because $54 < 117 < 136$, the null hypothesis is not rejected at the $\alpha = 0.10$ level of significance. There is no significant evidence of a difference between the unknown true medians of the measured values obtained by the two analytical methods.

9.2 TWO INDEPENDENT SAMPLES

This section presents nonparametric methods for comparing the locations and variances of two populations based on independent random samples selected from the populations. The corresponding parametric tests are the two sample t-test (Section 4.3.2) for comparing locations, and the F-test (Section 4.3.3) for comparing variances.

9.2.1 The Mann-Whitney Test for Comparing Locations

The Mann-Whitney test presented in this section is sometimes referred to as the Wilcoxon test or as the Wilcoxon rank-sum test. Mann, Whitney, and Wilcoxon each contributed to the development of this test; however, the Mann-Whitney name is used in this text to avoid confusion with the Wilcoxon signed ranks test.

The Mann-Whitney test assumes that random samples from two populations are available and that these samples were obtained independently from each other. Let x_1, x_2, \dots, x_n denote the random sample of size n from one population, and let y_1, y_2, \dots, y_m denote the random sample of size m from the other population. Let $N = n + m$. Assume that the measurement scale for the random variables X and Y is at least ordinal. It is of interest to test the hypothesis

$$H_0: \mu_X = \mu_Y$$

against the alternative

(9.9)

$$H_A: \mu_X \neq \mu_Y$$

where μ_X and μ_Y are the population means for the random variables X and Y, respectively. It is assumed that the population distribution functions of X and Y are the same except for a possible difference in means (locations). This implies equal variances; that is, $\sigma_X^2 = \sigma_Y^2$. This assumption should be investigated using the squared ranks test (Section 9.2.2).

The test is conducted by combining the samples from the two populations and ranking the combined sample using ranks 1 to $N = n + m$. If several sample values are exactly equal (tied), assign to each the average of the ranks that would otherwise have been assigned. Let $R(x_i)$ and $R(y_j)$ denote the ranks assigned to x_i and y_j for all i and j . The Mann-Whitney test statistic is the sum of the ranks assigned to the x_i 's; that is

$$W' = \sum_{i=1}^n R(x_i) \quad (9.10)$$

If there are more than a few ties, Conover (1980) presents an alternate test statistic. The calculated W' is compared to percentiles w_p obtained from Table A13. If W' is too large or too small, then H_0 in Equation 9.9 is rejected. More specifically, reject H_0 at the α level of significance if $W' > w_{1-\alpha/2}$ or if $W' < w_{\alpha/2}$. For the one-sided alternative $H_A: \mu_X > \mu_Y$, reject H_0 if $W' > w_{1-\alpha}$. For the one-sided alternative $H_A: \mu_X < \mu_Y$, reject H_0 if $W' < w_\alpha$.

Example 9.6 A facility producing UO_2 powder must check whether the percent uranium factor for two batches of powder is the same. Eight samples from each of the two batches are drawn and the percent uranium determined. The data are displayed in Table 9.3. Use the Mann-Whitney test with the $\alpha = 0.10$ level of significance.

The validity of the assumptions underlying the Mann-Whitney test should be checked. The samples from each batch were randomly and

†The ranks assigned to the y_j could also be used, because the representation of the populations by X and Y is arbitrary.

TABLE 9.3
Percent Uranium in UO₂

| Batch 1 | Batch 2 |
|---------|---------|
| 86.4 | 86.2 |
| 86.8 | 85.8 |
| 87.5 | 86.1 |
| 88.2 | 84.8 |
| 86.0 | 87.1 |
| 86.9 | 86.5 |
| 87.7 | 85.7 |
| 86.8 | 86.7 |

independently collected. Application of the runs test (Section 9.1.1) to each sample indicated no significant nonrandomness. The assumption of equal variances is substantiated in Example 9.7.

The Mann-Whitney test is performed by ranking the combined samples. This process is easier if the individual samples are ordered first. The ranks are given in Table 9.4.

TABLE 9.4
Ranks of Combined Samples from Batches 1 and 2

| Batch 1 | Combined rank | Batch 2 | Combined rank |
|---------|---------------|---------|---------------|
| 86.0 | 4 | 84.8 | 1 |
| 86.4 | 7 | 85.7 | 2 |
| 86.8 | 10.5 | 85.8 | 3 |
| 86.8 | 10.5 | 86.1 | 5 |
| 86.9 | 12 | 86.2 | 6 |
| 87.5 | 14 | 86.5 | 8 |
| 87.7 | 15 | 86.7 | 9 |
| 88.2 | 16 | 87.1 | 13 |

Because there is only one tie (two equal values), the test statistic is computed from Equation 9.10, which gives

$$W' = \sum_{i=1}^8 (\text{batch 1 ranks}) = 89$$

For $n = m = 8$ and $\alpha = 0.05$, the values $w_{0.05} = 50$ and $w_{0.095} = 86$ are

obtained from Table A13. Because $89 > 86$, H_0 is rejected, implying that the percent uranium factors for the two batches are different.

Recall that the parametric counterpart to the Mann-Whitney test is the two sample t-test (Section 4.3.2.). If the population distributions are actually normal, the t-test is more powerful than the Mann-Whitney test, although the difference in power is small. For many nonnormal distributions, the Mann-Whitney test is more powerful than the t-test (also note that the t-test is not strictly valid for nonnormal distributions). Thus, a convincing argument could be made for always using the Mann-Whitney test. It is certainly preferred if the distributions of X and Y are unknown or if there is any doubt about the normality assumption.

9.2.2 The Squared Ranks Test for Comparing Variances

Suppose that it is of interest to test the hypothesis

$$H_0: \sigma_X^2 = \sigma_Y^2$$

against the alternative

(9.11)

$$H_A: \sigma_X^2 \neq \sigma_Y^2$$

This section presents a method for testing H_0 based on independent samples collected from two populations. Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m denote independent random samples of size n and m from two populations. Let $N = n + m$. The measurement scale of X and Y must be at least interval.

The squared ranks test of H_0 was developed by Conover and Iman (1978) and is conducted as follows. Compute

$$u_i = |x_i - \mu_x|, \quad i = 1, 2, \dots, n$$

and

$$v_j = |y_j - \mu_y|, \quad j = 1, 2, \dots, m$$

where μ_X and μ_Y are the population means. If μ_X and μ_Y are unknown, substitute the sample means \bar{x} and \bar{y} (the test is still approximately valid). Then assign the ranks 1 to $N = n + m$ to the combined sample of u_i 's and v_j 's. If any values of u_i and/or v_j are tied, assign to each the average

of the ranks that would have been assigned had there been no ties. Let the resulting ranks be denoted by $R(u_i)$ and $R(v_j)$.

If there are no ties, the test statistic is the sum of the squared ranks assigned to the u_i ; that is,

$$T_X^* = \sum_{i=1}^n [R(u_i)]^2 \quad (9.12)$$

If there are ties, use the test statistic

$$T^* = \frac{T_X^* - n\bar{R}^2}{\left[\frac{nm}{N(N-1)} \sum_{i=1}^N R_i^4 - \frac{nm}{N-1} (\bar{R}^2)^2 \right]^{1/2}} \quad (9.13)$$

where

$$\bar{R}^2 = \frac{1}{N} \left\{ \sum_{i=1}^n [R(u_i)]^2 + \sum_{j=1}^m [R(v_j)]^2 \right\}$$

and

$$\sum_{i=1}^N R_i^4 = \sum_{i=1}^n [R(u_i)]^4 + \sum_{j=1}^m [R(v_j)]^4$$

Exact percentiles w_p of the distribution of T_X^* when no ties are present are given in Table A14. The distribution of T^* when ties are present is approximated by the standard normal distribution, and w_p is approximated by z_p from Table A3. The two-sided test of H_0 in Equation 9.11 is performed by rejecting H_0 at the significance level α when T_X^* (or T^*) is less than $w_{\alpha/2}$ or greater than $w_{1-\alpha/2}$. For the one-sided hypothesis $H_A: \sigma_X^2 > \sigma_Y^2$, reject H_0 if T_X^* (or T^*) $> w_{1-\alpha}$. For the one-sided hypothesis $H_A: \sigma_X^2 < \sigma_Y^2$, reject H_0 if T_X^* (or T^*) $< w_\alpha$.

Example 9.7 Apply the squared ranks test to the data in Example 9.6 to test the hypothesis of equal variances. Use the $\alpha = 0.10$ level of significance.

Because the population means are unknown, the sample estimates $\bar{x} = 87.04$ and $\bar{y} = 86.11$ are used. The absolute values of the deviations and their ranks in the combined sample are displayed in Table 9.5.

TABLE 9.5
Deviations and Ranks

| $u_i = x_i - \bar{x} $ | $R(u_i)$ | $v_j = y_j - \bar{y} $ | $R(v_j)$ |
|-----------------------------------|----------|---------------------------------|----------|
| 1.04 | 14 | 1.31 | 16 |
| 0.64 | 11 | 0.41 | 8 |
| 0.24 | 4.5 | 0.31 | 6 |
| 0.24 | 4.5 | 0.01 | 1 |
| 0.14 | 3 | 0.09 | 2 |
| 0.46 | 9 | 0.39 | 7 |
| 0.66 | 12 | 0.59 | 10 |
| 1.16 | 15 | 0.99 | 13 |
| $\sum_{i=1}^8 [R(u_i)]^2 = 816.5$ | | $\sum_{j=1}^8 [R(v_j)]^2 = 679$ | |

Because there are ties in the data, Equation 9.13 is applied to calculate the test statistic

$$T^* = \frac{816.5 - 8(816.5 + 679)/16}{\left[\frac{(8)(8)}{(16)(15)(243,787.13)} - \frac{(8)(8)}{15} \left(\frac{816.5 + 679}{16} \right)^2 \right]^{1/2}}$$

$$= \frac{68.75}{166.54} = 0.41$$

The values $w_{0.05} = -1.65$ and $w_{0.95} = 1.65$ are obtained from the standard normal distribution (Table A3). Because $-1.65 < 0.41 < 1.65$, H_0 is not rejected at the $\alpha = 0.10$ significance level. There is no significant evidence that the variances of the two populations are different.

Several other nonparametric tests for comparing the variances of two populations exist. Three of these, the Siegel-Tukey test, the Freund-Ansari test, and the Mood test are discussed by Bradley (1968) and briefly in Conover (1971). The squared ranks test discussed here is more powerful than the Siegel-Tukey and Freund-Ansari tests and is quite similar to the Mood test.

The parametric counterpart to the squared ranks test is the F-test (Section 4.3.3). When the x_i and y_j are actually from normal distributions, the F-test is preferred. Because the F-test is quite sensitive to departures from the assumption of normality, the squared ranks test is preferred when the distributions of X and Y are unknown or when the assumption of normality is in doubt.

9.3 SEVERAL INDEPENDENT SAMPLES

In this section nonparametric tests are presented for comparing the locations and the variances of several populations based on independent samples. The nonparametric tests for comparing locations are the median test and the Kruskal-Wallis test. The parametric test corresponding to these tests is the one-way analysis of variance (Section 5.1). The nonparametric test for comparing variances is an extension of the squared ranks test from Section 9.2.2. Its parametric counterparts are not presented in this text.

9.3.1 Tests for Comparing Locations

Two nonparametric tests of location are considered in this section, the median test and the Kruskal-Wallis test. The Kruskal-Wallis test has more stringent assumptions than the median test and is based on the ranks of the observations.

9.3.1.1 The Median Test

The median test is designed to investigate whether several samples come from populations having the same median. Formally, it is of interest to test the hypothesis

$$H_0 : \text{All populations have the same median}$$

against the alternative

(9.14)

$$H_A : \text{At least two of the populations have different medians}$$

The median test assumes that the sample from each population is random and that these samples are collected independently of each other. The measurement scale for each random variable is assumed to be at least ordinal. It is also assumed that if H_0 is true, all populations have the same probability of an observation exceeding the grand median (the median of all the samples combined).

To conduct the median test, compute the grand median of the combined samples. Let O_{i1} be the number of observations in the i^{th} sample that

exceed the grand median and let O_{2i} be the number less than or equal to the grand median. Arrange the O_{1i} and O_{2i} , $i = 1, 2, \dots, k$ into a $2 \times k$ contingency table as follows:

| | | Sample | | | | |
|----------|--|----------|----------|-------|----------|--------|
| | | 1 | 2 | | k | Totals |
| > Median | | O_{11} | O_{12} | | O_{1k} | a |
| ≤ Median | | O_{21} | O_{22} | | O_{2k} | b |
| Totals | | n_1 | n_2 | | n_k | N |

The test statistic is given by

$$\chi^* = \frac{N^2}{ab} \sum_{i=1}^k \frac{O_{fi}^2}{n_i} - \frac{aN}{b} \tag{9.15}$$

The exact distribution of χ^* is difficult to tabulate, but it can be satisfactorily approximated by the chi-square distribution with $k - 1$ degrees of freedom. Reject H_0 at the approximate significance level α if $\chi^* > \chi_{1-\alpha}^2$ ($k - 1$), obtained from Table A4.

The chi-square approximation may not be very accurate if some of the n_i are small. Conover (1980) suggests that, in general, the approximation may not be satisfactory if more than 20% of the n_i 's are less than 10 or if any of the n_i 's are less than 2. An exception to this rule can be made for larger values of k ; if most of the n_i 's are about equal to each other, then values of n_i as small as 2 are allowed.

Example 9.8 A facility has four production lines that each create a liquid waste stream. All four of the waste streams go to the same storage area, where they are sampled and analyzed (uranium concentration in ppm) for accountability purposes. The facility must test whether the four waste streams are entering the storage area with the same uranium concentration. Random samples from the four waste streams are collected and analyzed for uranium concentrations. The results are displayed in Table 9.6. Use the median test to test the hypothesis that there are no differences in the uranium concentrations of waste from the four production lines. Use the $\alpha = 0.10$ level of significance.

To apply the median test, the grand median of the combined samples is found to be 23. Using this value, the following 2×4 contingency table is developed as outlined earlier.

| | Line | | | | Totals |
|---------------|------|---|---|---|--------|
| | 1 | 2 | 3 | 4 | |
| > 23 | 3 | 5 | 2 | 5 | 15 |
| ≤ 23 | 7 | 3 | 7 | 3 | 20 |
| Totals | 10 | 8 | 9 | 8 | 35 |

The test statistic is calculated from Equation 9.15, which gives

$$\begin{aligned} \chi^* &= \frac{(35)^2}{(15)(20)} \left(\frac{3^2}{10} + \frac{5^2}{8} + \frac{2^2}{9} + \frac{5^2}{8} \right) - \frac{35(15)}{20} \\ &= 31.01 - 26.25 = 4.76 \end{aligned}$$

TABLE 9.6
Uranium Concentration in ppm

| Line 1 | Line 2 | Line 3 | Line 4 |
|--------|--------|--------|--------|
| 13 | 34 | 17 | 33 |
| 23 | 16 | 26 | 30 |
| 23 | 20 | 11 | 23 |
| 14 | 35 | 13 | 24 |
| 25 | 17 | 24 | 32 |
| 22 | 30 | 22 | 20 |
| 35 | 27 | 8 | 17 |
| 13 | 28 | 20 | 32 |
| 24 | | 23 | |
| 20 | | | |

Because $k = 4$ and the n_i are approximately equal, the chi-square approximation is probably satisfactory. For $\alpha = 0.10$, the value $\chi_{0.90}^2(3) = 6.25$ is obtained from Table A4. Because $4.76 < 6.25$, H_0 is not rejected. There is no significant evidence that the median uranium concentrations for the four lines are different.

9.3.1.2 The Kruskal-Wallis Test

The Kruskal-Wallis test, like the median test, is used to compare the locations of several populations. It assumes that a random sample of observations from each population is available, that there is mutual indepen-

dence among the samples, and that the measurement scale is at least ordinal. A final assumption is that either the population distribution functions are identical or that only their means differ (this includes the assumption that the variances of the populations are identical).

Let the random sample from the i^{th} population be denoted by $x_{i1}, x_{i2}, \dots, x_{in_i}$, $i = 1, 2, \dots, k$. The total number of observations in all samples is obtained by summing the individual sample sizes; that is,

$$N = \sum_{i=1}^k n_i$$

Rank the combined samples (smallest to largest) from 1 to N . If several observations are equal (tied), assign to each the average of the ranks that would otherwise have been assigned. Let $R(x_{ij})$ represent the rank assigned to x_{ij} , and let R_i be the sum of the ranks assigned to the i^{th} sample; that is,

$$R_i = \sum_{j=1}^{n_i} R(x_{ij}) \quad i = 1, 2, \dots, k$$

To test the hypothesis

H_0 : All k population distribution functions are identical

against the alternative

(9.16)

H_A : At least two populations have unequal means

compute the test statistic

$$\chi' = \frac{1}{S^2} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \quad (9.17)$$

where

$$S^2 = \begin{cases} N(N+1)/12 & \text{when no ties are present} \\ \frac{1}{N-1} \left(\sum_{ij} [R(x_{ij})]^2 - \frac{N(N+1)^2}{4} \right) & \text{when ties are present} \end{cases} \quad (9.18)$$

While some tables of exact quantiles of the distribution of χ' do exist, they are rather limited. The chi-square distribution with $k - 1$ degrees of freedom (Table A4) provides a satisfactory approximation to the distribution of χ' . The null hypothesis H_0 in Equation 9.16 is rejected at significance level α if $\chi' > \chi_{1-\alpha}^2(k - 1)$.

Rejection of H_0 implies only that at least one population mean differs from at least one other. The following multiple comparisons procedure can be used to indicate which of the population means differ. It can be concluded that the means of populations i and j are different if

$$\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right| > t_{1-\alpha/2}(N-k) \left[S^2 \frac{N-1-\chi'}{N-k} \right]^{1/2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)^{1/2} \quad (9.19)$$

where $t_{1-\alpha/2}(N-k)$ is obtained from Table A5, χ' is computed from Equation 9.17, and S^2 is computed from Equation 9.18. The α value should be the same as that used in the Kruskal-Wallis test. This procedure is repeated for all pairs of populations.

Example 9.9 For the data in Example 9.8, compare the mean uranium concentrations of the four lines by using the Kruskal-Wallis test with $\alpha = 0.10$.

The results in Example 9.10 provide no significant evidence that the variances of the four batches differ. Thus, it is plausible that the stricter assumptions required for the Kruskal-Wallis test, as compared to the median test, are satisfied.

The ranks of the combined sample observations are displayed in Table 9.7, where the x_{ij} values within each sample are arranged in rank order. The test statistic is calculated from Equation 9.17, where $S^2 = 104.47$ is obtained from the second formula in Equation 9.18 because ties exist in the data. This gives

$$\chi' = \frac{1}{104.47} \left[\frac{(159.5)^2}{10} + \frac{(177.5)^2}{8} + \frac{(109.5)^2}{9} + \frac{(183.5)^2}{8} - \frac{35(36)^2}{4} \right] = 6.54$$

TABLE 9.7
Combined Sample Ranks of Uranium Concentrations

| Line 1 | | Line 2 | | Line 3 | | Line 4 | |
|---------------|------|---------------|------|---------------|------|---------------|------|
| x_{1j} | Rank | x_{2j} | Rank | x_{3j} | Rank | x_{4j} | Rank |
| 13 | 4 | 16 | 7 | 8 | 1 | 17 | 9 |
| 13 | 4 | 17 | 9 | 11 | 2 | 20 | 12.5 |
| 14 | 6 | 20 | 12.5 | 13 | 4 | 23 | 18.5 |
| 20 | 12.5 | 27 | 26 | 17 | 9 | 24 | 22 |
| 22 | 15.5 | 28 | 27 | 20 | 12.5 | 30 | 28.5 |
| 23 | 18.5 | 30 | 28.5 | 22 | 15.5 | 32 | 30.5 |
| 23 | 18.5 | 34 | 33 | 23 | 18.5 | 32 | 30.5 |
| 24 | 22 | 35 | 34.5 | 24 | 22 | 33 | 32 |
| 25 | 24 | | | 26 | 25 | | |
| 35 | 34.5 | | | | | | |
| $R_1 = 159.5$ | | $R_2 = 177.5$ | | $R_3 = 109.5$ | | $R_4 = 183.5$ | |

Because $6.54 > \chi_{0.90}^2(3) = 6.25$, the null hypothesis is rejected at the $\alpha = 0.10$ significance level. This implies that the unknown true mean uranium factor of at least one batch differs from that of at least one other. Note that the Kruskal-Wallis test has detected a difference, whereas the median test in Example 9.8 did not. This illustrates that the Kruskal-Wallis test can be more powerful than the median test when the more stringent assumptions are satisfied.

Because H_0 was rejected, the multiple comparisons procedure can be applied to indicate which of the means differ.

To apply the procedure, given by Equation 9.19, the quantity

$$\begin{aligned}
 t_{0.95}(N - k) \left[S^2 \frac{N - 1 - \chi'}{N - k} \right]^{1/2} &= (1.70) \left[(104.47) \frac{35 - 1 - 6.54}{35 - 4} \right]^{1/2} \\
 &= 16.35
 \end{aligned}$$

is computed. Then application of Equation 9.19 to all pairs (i, j) yields the following results:

| Populations | | $\left \frac{R_i}{n_i} - \frac{R_j}{n_j} \right $ | $16.35 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)^{1/2}$ | Significant
at $\alpha = 0.10$ |
|-------------|---|--|--|-----------------------------------|
| i | j | | | |
| 1 | 2 | 6.24 | 7.76 | No |
| 1 | 3 | 3.78 | 7.51 | No |
| 1 | 4 | 6.99 | 7.76 | No |
| 2 | 3 | 10.02 | 7.94 | Yes |
| 2 | 4 | 0.75 | 8.18 | No |
| 3 | 4 | 10.77 | 7.94 | Yes |

These results indicate that the unknown true mean uranium concentration for line 3 is lower than the true mean concentrations for lines 2 and 4.

Both the median test and the Kruskal-Wallis test for comparing the means of k populations have been considered. The Kruskal-Wallis test requires the stringent assumption that all populations have identical distribution functions, except for possibly different means (this includes the assumption that the population variances must be identical). If these more stringent assumptions are true, the Kruskal-Wallis test is more powerful for detecting population location differences than the median test because ranking the data retains more information than dichotomizing. However, if the populations have different distribution functions, other than different locations, then the Kruskal-Wallis test is not applicable and the median test should be applied.

9.3.2 The k -Sample Squared Ranks Test for Comparing Variances

Now consider a nonparametric test for comparing the variances of several populations. It is of interest to test the hypothesis

H_0 : All k populations have identical distributions except for possibly different means

against the alternative

(9.20)

H_A : At least two of the populations have unequal variances.

The k -sample squared ranks test is an extension of the two-sample squared ranks test presented in Section 9.2.2. Let $x_{i1}, x_{i2}, \dots, x_{in_i}$ denote a random sample of size n_i from the i^{th} population ($i = 1, 2, \dots, k$). Assume that the k random samples are independent and that the measurement scale is at least ordinal.

From each observation x_{ij} , subtract its population mean μ_i , or its sample mean \bar{x}_i if μ_i is unknown. Then combine all samples and rank the absolute differences $|x_{ij} - \mu_i|$ or $|x_{ij} - \bar{x}_i|$ from smallest to largest, assigning average ranks in case of ties. Denote the rank of $|x_{ij} - \mu_i|$ by $R(x_{ij})$. The test statistic is given by

$$T^* = \frac{1}{D^2} \left[\sum_{i=1}^k \frac{S_i^2}{n_i} - N(\bar{S})^2 \right] \quad (9.21)$$

where

n_i = the number of observations in sample i

$N = n_1 + n_2 + \dots + n_k$

$S_i = \sum_{j=1}^{n_i} [R(x_{ij})]^2$

$\bar{S} = \frac{1}{N} \sum_{i=1}^k S_i$

$D^2 = \frac{1}{N-1} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} [R(x_{ij})]^4 - N(\bar{S})^2 \right]$

If there are no ties, then D^2 and \bar{S} simplify to

$$D^2 = N(N+1)(2N+1)(8N+11)/180$$

and

$$\bar{S} = (N+1)(2N+1)/6$$

The null hypothesis is rejected at the α level of significance if T^* exceeds $\chi_{1-\alpha}^2(k-1)$, obtained from Table A4. Rejection of H_0 indicates that at least two of the populations have unequal variances, but it gives no indication of which variances differ. Multiple comparisons among the k populations are made by comparing two variances at a time. It can be concluded that the variances of populations i and j differ if

$$\left| \frac{S_i}{n_i} - \frac{S_j}{n_j} \right| > t_{1-\alpha/2}(N-k) \left(D^2 \frac{N-1-T^*}{N-k} \right)^{1/2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)^{1/2} \quad (9.22)$$

where $t_{1-\alpha/2}(N-k)$ is obtained from Table A5.

Example 9.10 By using the k -sample squared ranks test and the data from Example 9.8, test the hypothesis that the variances of the uranium concentrations of the four production lines are equal. Use the $\alpha = 0.10$ level of significance.

The $|x_{ij} - \bar{x}_i|$ are computed and ranked as shown below. The sum of the squared ranks S_i is then computed for each production line.

| Line 1 | | Line 2 | | Line 3 | | Line 4 | |
|------------------------|------|------------------------|------|------------------------|------|------------------------|------|
| $ x_{1j} - \bar{x}_1 $ | Rank | $ x_{2j} - \bar{x}_2 $ | Rank | $ x_{3j} - \bar{x}_3 $ | Rank | $ x_{4j} - \bar{x}_4 $ | Rank |
| 8.20 | 28.5 | 8.12 | 27 | 1.22 | 4 | 6.62 | 23 |
| 1.80 | 6.5 | 9.88 | 33 | 7.78 | 26 | 3.62 | 12 |
| 1.80 | 6.5 | 5.88 | 21 | 7.22 | 25 | 3.38 | 11 |
| 7.20 | 24 | 9.12 | 31 | 5.22 | 17 | 2.38 | 9 |
| 3.80 | 14 | 8.88 | 30 | 5.78 | 20 | 5.62 | 18.5 |
| 0.80 | 1 | 4.12 | 15 | 3.78 | 13 | 6.38 | 22 |
| 13.80 | 35 | 1.12 | 2 | 10.22 | 34 | 9.38 | 32 |
| 8.20 | 28.5 | 2.12 | 8 | 1.78 | 5 | 5.62 | 18.5 |
| 2.80 | 10 | | | 4.78 | 16 | | |
| 1.20 | 3 | | | | | | |
| $S_1 = 3816.0$ | | $S_2 = 4413.0$ | | $S_3 = 3612.0$ | | $S_4 = 3067.5$ | |

The test statistic is calculated from Equation 9.21, where $\bar{S} = 425.96$, $D^2 = 144,556.49$, and $N = 35$. This gives

$$T^* = \frac{1}{144,556.49} \left[\frac{(3816.0)^2}{10} + \frac{(4413.0)^2}{8} + \frac{(3612.0)^2}{9} + \frac{(3067.5)^2}{8} - 35(425.96)^2 \right] = 1.15$$

Because 1.15 is less than $\chi_{0.90}^2(3) = 6.25$, H_0 is not rejected. There is no significant indication that the uranium concentration variances for the four production lines are different.

9.4 SEVERAL RELATED SAMPLES

In this section, nonparametric tests are presented for comparing the locations of several populations using related samples. These tests are extensions of the matched pairs tests presented in Section 9.1.3. The corresponding parametric method is the two-way analysis of variance using data from a randomized complete block design (Section 7.2.2), which is an extension of the paired t -test (Section 4.4).

9.4.1 Comparison of Locations for Randomized Complete Block Designs

Randomized complete block designs, including the concept of blocking in experimentation, are discussed in Section 7.2.2. Briefly, such designs compare the means resulting from $k \geq 2$ different treatments, where the observations are arranged in blocks. The blocks for these designs are groups of k experimental units where the units within a group are similar, but significant differences among the groups are expected in the experimental data. An example of such a design in nuclear material control is where identical samples from each of m items of SNM are sent to each of k laboratories for analysis. The laboratories are regarded as the treatments while the m items are the blocks. This type of application is illustrated in Example 9.11.

It is of interest to test the hypothesis

$$H_0 : \text{There are no differences among the population locations} \\ \text{corresponding to the } k \text{ treatment groups}$$

against the alternative

(9.23)

$$H_A : \text{At least one of the } k \text{ treatments tends to yield larger values} \\ \text{than at least one other.}$$

Three nonparametric tests are presented for testing this hypothesis using related samples data: the Quade test, the Friedman test, and the rank ANOVA test.

9.4.1.1 The Quade Test

The Quade test is an extension of the matched pairs application of the Wilcoxon signed ranks test. The major assumption is the independence of the blocks; the results within one block do not influence the results within other blocks.

The Quade test is conducted as follows. Let x_{ij} ($i = 1, 2, \dots, b$; $j = 1, 2, \dots, k$) be the observation from the i^{th} block for the j^{th} treatment in a randomized complete block design. Within each block, rank the observations $x_{i1}, x_{i2}, \dots, x_{ik}$ from smallest to largest by assigning rank 1 to the smallest value and k to the largest. In case of ties, assign average ranks. Let $R(x_{ij})$ be the rank assigned to x_{ij} .

Next, calculate the sample range within each block, i.e.,

$$\text{Range}_i = \max_j (x_{ij}) - \min_j (x_{ij})$$

Then rank the blocks from 1 to b according to the Range_i values. Use average ranks in case of ties. Let Q_i be the rank of the i^{th} block.

The test statistic is given by

$$F^* = \frac{(b-1)B_1}{A_1 - B_1} \quad (9.24)$$

where

$$A_1 = \sum_{i=1}^b \sum_{j=1}^k S_{ij}^2$$

$$B_1 = \frac{1}{b} \sum_{j=1}^k S_j^2$$

$$S_{ij} = Q_i \left[R(x_{ij}) - \frac{k+1}{2} \right]$$

$$S_j = \sum_{i=1}^b S_{ij}$$

The null hypothesis is rejected at the α level of significance if F^* exceeds $F_{1-\alpha}[k-1, (b-1)(k-1)]$, obtained from Table A6. The F-distribution approximates the exact distribution of F^* with the approximation improving as the number of blocks b increases.

If the null hypothesis is rejected, pairwise differences are examined by using a multiple comparisons procedure. It can be concluded that the locations resulting from treatments j and l are different if

$$|S_j - S_l| > t_{1-\alpha/2}[(b-1)(k-1)] \left[\frac{2b(A_1 - B_1)}{(b-1)(k-1)} \right]^{1/2} \quad (9.25)$$

where S_j, S_l, A_1 , and B_1 are as given above, and $t_{1-\alpha/2}[(b-1)(k-1)]$ is obtained from Table A5. Note, however, that this multiple comparisons procedure should only be applied if the overall null hypothesis is rejected.

Example 9.11 Three laboratories are each given a sample of PuO_2 from each of 10 containers and are asked to measure the percent plutonium by weight in each sample. The results are displayed in Table 9.8. Treat the containers as blocks, and use the Quade test to test the hypothesis that there are no biases among the three laboratories. Use the $\alpha = 0.05$ level of significance.

TABLE 9.8
Percent Pu by Weight

| Container | Lab 1 | Lab 2 | Lab 3 |
|-----------|-------|-------|-------|
| 1 | 87.67 | 87.88 | 87.72 |
| 2 | 87.64 | 87.99 | 87.80 |
| 3 | 88.03 | 87.98 | 87.97 |
| 4 | 87.50 | 87.84 | 87.65 |
| 5 | 87.86 | 87.77 | 87.75 |
| 6 | 87.71 | 87.93 | 87.76 |
| 7 | 87.63 | 87.83 | 87.79 |
| 8 | 87.83 | 87.94 | 87.93 |
| 9 | 88.04 | 88.06 | 88.05 |
| 10 | 87.72 | 88.03 | 87.68 |

There is no reason to believe that the results for samples from any single container have affected the results for any other container; hence, the assumptions for the Quade test appear to be reasonable. The results of ranking the observations within blocks (containers) and then ranking the blocks are summarized in Table 9.9a. The calculated values of the S_{ij} and S_j are summarized in Table 9.9b.

Substituting the results from Table 9.9b into the equations for A_1 and B_1 yields

$$A_1 = (-6)^2 + (-9.5)^2 + (2)^2 + \dots + (-9.5)^2 = 768.0$$

$$B_1 = \frac{1}{10} [(-34.5)^2 + (49.5)^2 + (-15.0)^2] = 386.55$$

Application of Equation 9.24 yields the test statistic value

$$F^* = \frac{(b - 1) B_1}{A_1 - B_1} = \frac{(10 - 1)(386.55)}{768.0 - 386.55} = 9.12$$

TABLE 9.9a
 Quade Test Calculations: Ranking Results

| Container
(i) | Within container ranks
$R(x_{ij})$ | | | Range _i | Range, rank
Q_i |
|------------------|---------------------------------------|-------|-------|--------------------|----------------------|
| | Lab 1 | Lab 2 | Lab 3 | | |
| 1 | 1 | 3 | 2 | 0.21 | 6 |
| 2 | 1 | 3 | 2 | 0.35 | 9.5 |
| 3 | 3 | 2 | 1 | 0.06 | 2 |
| 4 | 1 | 3 | 2 | 0.34 | 8 |
| 5 | 3 | 2 | 1 | 0.11 | 3.5 |
| 6 | 1 | 3 | 2 | 0.22 | 7 |
| 7 | 1 | 3 | 2 | 0.20 | 5 |
| 8 | 1 | 3 | 2 | 0.11 | 3.5 |
| 9 | 1 | 3 | 2 | 0.02 | 1 |
| 10 | 2 | 3 | 1 | 0.35 | 9.5 |

TABLE 9.9b
 Quade Test Calculations: S_i and S_j Values

| Container
(i) | S_i | | |
|------------------|-------|-------|-------|
| | Lab 1 | Lab 2 | Lab 3 |
| 1 | -6 | 6 | 0 |
| 2 | -9.5 | 9.5 | 0 |
| 3 | 2 | 0 | -2 |
| 4 | -8 | 8 | 0 |
| 5 | 3.5 | 0 | -3.5 |
| 6 | -7 | 7 | 0 |
| 7 | -5 | 5 | 0 |
| 8 | -3.5 | 3.5 | 0 |
| 9 | -1 | 1 | 0 |
| 10 | 0 | 9.5 | -9.5 |
| S_j | -34.5 | 49.5 | -15.0 |

Table A6 does not give a value for $F_{0.95}(2,18)$. However, the values $F_{0.95}(2,15) = 3.68$ and $F_{0.95}(2,20) = 3.49$ are given and can be used with linear interpolation to compute the approximate value $F_{0.95}(2,18) \cong 3.49 + (3.68 - 3.49) [(20 - 18)/(20 - 15)] = 3.57$. Because $9.12 > 3.57$, the null hypothesis of no biases among the three laboratories is rejected. It should be noted that because $F^* = 9.12$ exceeds both $F_{0.95}(2,15)$ and $F_{0.95}(2,20)$, the interpolation is not necessary in order to

make the decision to reject H_0 . However, the interpolation was carried out for illustration. Application of the multiple comparisons procedure, given in Equation 9.25, indicates that Lab 2 tends to produce larger percent Pu than Labs 1 and 3. The results are displayed in the following table.

| j | l | $ S_j - S_l $ | $t_{0.975}(18)$ | $\left[\frac{2b(A_1 - B_1)}{(b-1)(k-1)} \right]^{1/2}$ | Significant
at $\alpha=0.05$ |
|-----|-----|---------------|-----------------|---|---------------------------------|
| 1 | 2 | 84.0 | | 43.23 | Yes |
| 1 | 3 | 19.5 | | 43.23 | No |
| 2 | 3 | 64.5 | | 43.23 | Yes |

9.4.1.2 The Friedman Test

The Friedman test is an extension of the matched pairs application of the sign test. The major assumption required for this test is the same as that for the Quade test (independence of blocks). Conover (1980) notes that when there are six or more treatments, the Friedman test is more powerful and is preferred over the Quade test.

The Friedman test of the hypothesis given in Equation 9.23 is conducted as follows. Let x_{ij} ($i = 1, 2, \dots, b$; $j = 1, 2, \dots, k$) be the observation from the i^{th} block for the j^{th} treatment in a randomized complete block design. Then let

$$R_j = \sum_{i=1}^b R(x_{ij})$$

where the $R(x_{ij})$ are the ranks within the blocks (found in the same manner as described in the previous section). The test statistic is given by

$$F' = \frac{(b-1)[B_2 - bk(k+1)^2/4]}{A_2 - B_2} \quad (9.26)$$

where

$$A_2 = \sum_{i=1}^b \sum_{j=1}^k [R(x_{ij})]^2$$

and

$$B_2 = \frac{1}{b} \sum_{j=1}^k R_j^2$$

The null hypothesis is rejected at the α level of significance if F' exceeds $F_{1-\alpha}[k-1, (b-1)(k-1)]$, obtained from Table A6. Note that the F -distribution only approximates the exact distribution of F' with the approximation improving as the number of blocks b increases.

If the null hypothesis is rejected, pairwise differences are examined by applying a multiple comparisons procedure. It can be concluded that treatments j and ℓ have resulted in different locations if

$$|R_j - R_\ell| > t_{1-\alpha/2}[(b-1)(k-1)] \left[\frac{2b(A_2 - B_2)}{(b-1)(k-1)} \right]^{1/2} \quad (9.27)$$

where R_j , R_ℓ , A_2 , and B_2 are as given above, and $t_{1-\alpha/2}[(b-1)(k-1)]$ is obtained from Table A5.

Example 9.12 Rework Example 9.11 using the Friedman test. The within block ranks $R(x_{ij})$ are displayed in Table 9.9a of Example 9.11. From these, the following quantities are computed:

$$R_1 = \sum_{i=1}^{10} R_{i1} = 1 + 1 + 3 + \dots + 1 + 2 = 15$$

$$R_2 = \sum_{i=1}^{10} R_{i2} = 3 + 3 + 2 + \dots + 3 + 3 = 28$$

$$R_3 = \sum_{i=1}^{10} R_{i3} = 2 + 2 + 1 + \dots + 2 + 1 = 17$$

$$A_2 = \sum_{i=1}^{10} \sum_{j=1}^3 [R(x_{ij})]^2 = 1^2 + 3^2 + 2^2 + 1^2 + \dots + 3^2 + 1^2 = 140$$

$$B_2 = \frac{1}{10} \sum_{j=1}^3 R_j^2 = \frac{1}{10} [(15)^2 + (28)^2 + (17)^2] = 129.8$$

The test statistic is computed from Equation 9.26, which gives

$$F' = \frac{(10 - 1) \left[129.8 - \frac{10(3)(4)^2}{4} \right]}{140 - 129.8} = 8.65$$

The value $F_{0.95}(2,18) \cong 3.57$ is obtained from Table A6 by interpolation (Example 9.11). Because $8.65 > 3.57$, H_0 is rejected at the $\alpha = 0.05$ level of significance. Application of the multiple comparisons procedure, given in Equation 9.27, indicates that Lab 2 tends to produce larger percent Pu values than Labs 1 and 3. The results are displayed in the following table. These conclusions are identical to those reached in Example 9.11 where the Quade test was used.

| j | q | $ R_j - R_q $ | $t_{0.975}(18) \left[\frac{20(A_2 - B_2)}{18} \right]^{1/2}$ | Significant
at $\alpha = 0.05$ |
|-----|-----|---------------|---|-----------------------------------|
| 1 | 2 | 13 | 7.07 | Yes |
| 1 | 3 | 2 | 7.07 | No |
| 2 | 3 | 11 | 7.07 | Yes |

9.4.1.3 The Rank ANOVA Test

The rank ANOVA test is the parametric two-way ANOVA test (Section 5.4), where the observations from a randomized complete block design (Section 7.2.2) are replaced by their ranks 1, 2, ..., bk. Tied observations are replaced by their average ranks. Iman, Hora, and Conover (1981) and Hora and Conover (1981) compare the robustness and power of the rank ANOVA test with the parametric ANOVA, the Quade test, and the Friedman test. The rank ANOVA test is more powerful than all of the other methods for many situations. When the underlying distributions are normal, the preferred method is the parametric ANOVA test performed on the original data rather than the ranks. When the underlying distribution is uniform, either the Quade test or the Friedman test should be used.

Example 9.13 Rework Example 9.11 using the rank ANOVA test with the $\alpha = 0.05$ level of significance.

The original data are displayed in Table 9.8 of Example 9.11. The ranks of all observations are displayed in Table 9.10.

TABLE 9.10
Ranks of Percent Pu Data

| Container | Lab 1 | Lab 2 | Lab 3 |
|-----------|-------|-------|-------|
| 1 | 5 | 19 | 8.5 |
| 2 | 3 | 25 | 14 |
| 3 | 26.5 | 24 | 23 |
| 4 | 1 | 17 | 4 |
| 5 | 18 | 12 | 10 |
| 6 | 7 | 20.5 | 11 |
| 7 | 2 | 15.5 | 13 |
| 8 | 15.5 | 22 | 20.5 |
| 9 | 28 | 30 | 29 |
| 10 | 8.5 | 26.5 | 6 |

Performing a two-way ANOVA using the ranks in Table 9.10 as data (see Section 5.4 for computational formulas) yields the following ANOVA table.

| Source | df | SS | MS | F |
|---------------------------|----|--------|-------|------|
| Containers (blocks) | 9 | 1236.8 | 137.4 | |
| Laboratories (treatments) | 2 | 508.9 | 254.4 | 9.15 |
| Error | 18 | 499.8 | 27.8 | |
| Total | 29 | 2245.5 | | |

The test statistic for comparing laboratories is $F = 9.15$, which is compared with $F_{0.95}(2,18) = 3.57$ obtained from Table A6 by interpolation. Because $9.15 > 3.57$, the null hypothesis is rejected at the $\alpha = 0.05$ level of significance. This agrees with the results of Examples 9.11 and 9.12.

9.4.2 Comparison of Locations for Other Designs

Chapter 7 introduced several different types of experimental designs. Conover (1980) presents the nonparametric Durbin test for comparing treatments using data from a balanced incomplete block design. Crouse (1968) and McDonald (1971) discuss nonparametric methods for analyz-

ing data from a 2^n factorial design. Tests for other specific designs are also found in the literature. However, there are many experimental designs for which specific nonparametric tests have not yet been developed.

Conover (1980) suggests a general approach for any experimental design. This approach involves replacing the observations by their ranks and then applying the parametric ANOVA technique appropriate for the particular design [see Conover (1980) for details].

9.5 MEASURES OF RANK CORRELATION

This section presents the concept of nonparametric correlation. The parametric analog of this concept is presented in Section 6.1.1, where the sample correlation coefficient $\hat{\rho}_{xy}$ is introduced as a measure of the strength of a linear relationship between two random variables X and Y. If one or more of the observed data pairs (x, y) is an outlier, $\hat{\rho}_{xy}$ may provide a misleading measure of the strength of the relationship. The impact of outliers is greatly reduced if the measure of correlation is based on ranks of the data. Two measures of **rank correlation**, Spearman's rho and Kendall's tau, are presented in this section.

In replacing data by ranks, some information may be lost, such as when interval or ratio scale data are reduced to the ordinal scale by the rank transformation. Because of this potential loss, measures of rank correlation are not thought of as measuring the strength of a linear relationship, but as measuring the strength of dependence in general. Thus, measures such as Spearman's rho and Kendall's tau are often used to test for statistical independence of the random variables X and Y. However, Spearman's rho and Kendall's tau are insensitive to some types of dependence. Thus, the use of these measures in testing for independence should be augmented by graphical techniques.

Although Spearman's rho and Kendall's tau can be used as alternate measures of correlation when the data are measured on the interval or ratio scales, they are the only applicable measures when either or both of the variables is measured on an ordinal scale.

9.5.1 Spearman's Rho

Assume that the data consist of a bivariate random sample of size n , denoted by $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Let $R(x_i)$ be the rank of x_i compared with the other observed X values, with $R(y_i)$ similarly defined for the observed Y values. In case of ties, assign the average of the ranks that would have been assigned had there been no ties. This ranking procedure assumes that the measurement scale for both X and Y is at least ordinal.

If there are no ties in the data for either X or Y, the measure of rank correlation known as Spearman's rho is given by

$$\tilde{\rho} = 1 - \frac{6 \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n(n^2 - 1)} \quad (9.28)$$

If there are ties in the data, $\tilde{\rho}$ is given by

$$\tilde{\rho} = \frac{\sum_{i=1}^n R(x_i) R(y_i) - n \left(\frac{n+1}{2} \right)^2}{\left[\sum_{i=1}^n [R(x_i)]^2 - n \left(\frac{n+1}{2} \right)^2 \right]^{1/2} \left[\sum_{i=1}^n [R(y_i)]^2 - n \left(\frac{n+1}{2} \right)^2 \right]^{1/2}} \quad (9.29)$$

Note that Equation 9.29 is just Equation 6.2 where the ranks $R(x_i)$ and $R(y_i)$ replace the x_i and y_i values. Thus, $-1 \leq \tilde{\rho} \leq 1$.

It is of interest to test the hypothesis

H_0 : There is no relationship between X and Y

against the alternative

(9.30)

H_A : There is either a tendency for larger values of X to be paired with larger values of Y or larger values of X to be paired with smaller values of Y.

H_0 is rejected if $\tilde{\rho} > w_{1-\alpha/2}$ or if $\tilde{\rho} < w_{\alpha/2}$, where w_p is found in Table A15 for $n \leq 30$. For larger n , an approximation is used as indicated in the footnote to Table A15. For the one-sided alternative, H_A : larger values of X tend to be paired with larger values of Y, reject H_0 if $\tilde{\rho} > w_{1-\alpha}$. For the one-sided alternative, H_A : larger values of X tend to be paired with smaller values of Y, reject H_0 if $\tilde{\rho} < w_{\alpha}$.

Example 9.14 Two uranium-bearing solid waste barrel standards are analyzed (counted) on the same day using the same NDA instrument on a semi-monthly basis. If there are shifts over time in the bias of the counter, the measurements (counts) from the two barrels should be highly correlated. The data displayed in Table 9.11 were generated over a 6-month period. Test the hypothesis of no relationship between the measurements of the two barrels by using Spearman's rho with the $\alpha = 0.05$ level of significance.

To compute Spearman's rho, the measurement values for each barrel are ranked. The ranks for both barrels are displayed in Table 9.12.

TABLE 9.11
NDA Uranium Measurements for
Two Waste Barrel Standards

| Time of observation | Barrel 1, g ²³⁵ U | Barrel 2, g ²³⁵ U |
|---------------------|------------------------------|------------------------------|
| 1 | 12.92 | 9.76 |
| 2 | 12.50 | 10.12 |
| 3 | 12.61 | 10.57 |
| 4 | 12.48 | 10.80 |
| 5 | 12.44 | 9.94 |
| 6 | 12.62 | 10.17 |
| 7 | 12.35 | 10.08 |
| 8 | 12.54 | 9.97 |
| 9 | 12.21 | 10.11 |
| 10 | 12.22 | 10.22 |
| 11 | 12.17 | 9.79 |
| 12 | 12.34 | 9.96 |

TABLE 9.12
Ranks of Uranium Measurements

| Time of observation | Ranks of U contents | |
|---------------------|---------------------|----------|
| | Barrel 1 | Barrel 2 |
| 1 | 12 | 1 |
| 2 | 8 | 8 |
| 3 | 10 | 11 |
| 4 | 7 | 12 |
| 5 | 6 | 3 |
| 6 | 11 | 9 |
| 7 | 5 | 6 |
| 8 | 9 | 5 |
| 9 | 2 | 7 |
| 10 | 3 | 10 |
| 11 | 1 | 2 |
| 12 | 4 | 4 |

Because there are no ties in the data, Spearman's rho is computed from Equation 9.28, which gives

$$\begin{aligned}\tilde{\rho} &= 1 - \frac{6[(12-1)^2 + (8-8)^2 + \dots + (4-4)^2]}{12(12^2-1)} \\ &= 1 - \frac{6(252)}{12(143)} = 1 - 0.88 = 0.12\end{aligned}$$

For $\alpha = 0.05$ and $n = 12$, the values $w_{0.975} = 0.5804$ and $w_{0.025} = -0.5804$ are obtained from Table A15. Because $-0.5804 < 0.12 < 0.5804$, H_0 is not rejected. There is no indication of a dependence in the counts from the two barrels; thus, there is no significant evidence of shifts over time in the bias of the counter.

9.5.2 Kendall's Tau

As with Spearman's rho, assume that the data consist of a bivariate random sample of size n , denoted by $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Two observations, say (x_1, y_1) and (x_2, y_2) , are **concordant** if both members of one observation are larger than their respective members of the other observation (i.e., $x_1 > x_2$ and $y_1 > y_2$, or $x_1 < x_2$ and $y_1 < y_2$). Let N_c represent the number of concordant pairs out of the $\binom{n}{2}$ possible pairs of bivariate observations. A pair of observations (x_1, y_1) and (x_2, y_2) is **discordant** if $x_1 > x_2$ and $y_1 < y_2$, or if $x_1 < x_2$ and $y_1 > y_2$. Let N_d represent the number of discordant pairs. If $x_1 = x_2$ or $y_1 = y_2$, the pair of points is neither concordant nor discordant. The identification of the number of concordant and discordant pairs is simplified if the data points (x_i, y_i) are sorted according to increasing values of X .

Kendall's tau is calculated as

$$\tau = \frac{N_c - N_d}{n(n-1)/2} \quad (9.31)$$

As with the previously defined measures of correlation, $-1 \leq \tau \leq 1$. The extreme values of -1 or 1 are reached when all pairs are discordant or concordant, respectively.

Kendall's tau may be used to test the hypothesis in Equation 9.30. Some arithmetic is saved, however, by only calculating the numerator portion of Equation 9.31; that is,

$$T^* = N_c - N_d \quad (9.32)$$

Quantiles of the distribution of T^* are given in Table A16. The two-sided test is performed by rejecting H_0 at the α level of significance if $T^* > w_{1-\alpha/2}$ or if $T^* < w_{\alpha/2}$. For the one-sided alternative H_A : larger values of X tend to be paired with larger values of Y , reject H_0 if $T^* > w_{1-\alpha}$. For the one-sided alternative H_A : larger values of X tend to be paired with smaller values of Y , reject H_0 if $T^* < w_\alpha$.

The following example illustrates the procedures for calculating and testing hypotheses with Kendall's tau.

Example 9.15 Test the hypothesis indicated in Example 9.14 by using Kendall's tau with the $\alpha = 0.05$ level of significance.

The data pairs are sorted by the x_i values and the number of concordant and discordant pairs are determined. The results are displayed in Table 9.13. As a check, $N_c + N_d + (\text{number of pairs with ties})$ should equal $n(n - 1)/2$. Because there are no ties, note that $N_c + N_d = n(n - 1)/2 = 66$.

To illustrate the calculations summarized in Table 9.13, consider the point (12.50, 10.12). Because the points in Table 9.13 are arranged with the x_i 's in increasing order, only the points below (12.50, 10.12) in the table are considered in attempting to identify concordant pairs. The concordance or discordance of (12.50, 10.12) with points above it in the table was evaluated when the procedure was applied to those points (Table 9.13). The points (12.61, 10.57) and (12.62, 10.17) form concordant pairs with (12.50, 10.12). Points that form discordant pairs with (12.50, 10.12) are those below it in Table 9.13, which have y_i values smaller than 10.12. The points (12.54, 9.97) and (12.92, 9.76) form discordant pairs with (12.50, 10.12). The results of applying this counting procedure are summarized in Table 9.13.

TABLE 9.13
Calculation of Kendall's Tau

| (x_i, y_i) | Number of
concordant pairs
involving (x_i, y_i) | Number of
discordant pairs
involving (x_i, y_i) |
|----------------|---|---|
| (12.17, 9.79) | 10 | 1 |
| (12.21, 10.11) | 5 | 5 |
| (12.22, 10.22) | 2 | 7 |
| (12.34, 9.96) | 6 | 2 |
| (12.35, 10.08) | 4 | 3 |
| (12.44, 9.94) | 5 | 1 |
| (12.48, 10.80) | 0 | 5 |
| (12.50, 10.12) | 2 | 2 |
| (12.54, 9.97) | 2 | 1 |
| (12.61, 10.57) | 0 | 2 |
| (12.62, 10.17) | 0 | 1 |
| (12.92, 9.76) | 0 | 0 |
| | $N_c = 36$ | $N_d = 30$ |

Kendall's tau is calculated from Equation 9.31, which gives

$$\tau = \frac{N_c - N_d}{n(n-1)/2} = \frac{36 - 30}{12(11)/2} = \frac{6}{66} = 0.09$$

For $\alpha = 0.05$ and $n = 12$, the values $w_{0.025} = -28$ and $w_{0.975} = 28$ are obtained from Table A16. Because $T^* = N_c - N_d = 6$ falls between -28 and 28 , the null hypothesis is not rejected at the $\alpha = 0.05$ level of significance. This is consistent with the results of Example 9.14.

9.6 GOODNESS-OF-FIT TESTS

As noted in the introductory comments of this chapter, the distinction between parametric and nonparametric procedures is that parametric procedures require specific assumptions about the form of probability distributions from which the empirical observations arise. If these assumptions are valid, the parametric techniques are preferred over the nonparametric techniques. However, it is never known for certain whether or not assumptions about a probability distribution are valid. In some situations it is possible to apply statistical hypothesis testing techniques to investigate the validity of an assumption. Hypothesis tests about the form of a probability distribution are often referred to as **goodness-of-fit tests**.

There are two major categories of goodness-of-fit tests. Tests in the first category compare observations from an unknown distribution with the expected outcome of sampling from a particular distribution that has been **completely specified**. The phrase "completely specified" means that both the form and any parameters of the reference distribution are specified without using the observations from the unknown distribution to estimate them. Two goodness-of-fit tests of this type, the **chi-square test** and the **Kolmogorov test** are presented in Sections 9.6.1 and 9.6.2, respectively.

Goodness-of-fit tests from the other major category compare observations from an unknown distribution with the expected outcome of sampling from a family of distributions; a particular member of the family is chosen for comparison on the basis of the observations from the unknown distribution. For example, it may be of interest to test whether a certain set of observations is from a normal distribution. The specific normal distribution is determined by estimating its parameters (the mean and variance) using the observations from the unknown distribution. Tests of this type are the **Lilliefors test** and the **Shapiro-Wilk test**. These are presented in Section 9.6.3.

9.6.1 Chi-Square Test

The **chi-square test** can be used to investigate assumptions about both discrete and continuous distributions. Let $F(x)$ be the true cumulative dis-

tribution function (cdf) of a random variable X (Section 2.2), and let $F^*(x)$ be a **completely specified**† cdf. $F^*(x)$ is the hypothesized distribution for the test; it is of interest to test whether the true cdf $F(x)$ is the same as the hypothesized cdf $F^*(x)$. Formally stated, it is of interest to test the hypothesis

$$H_0: F(x) = F^*(x), \text{ for all } x$$

against the alternative (9.33)

$$H_A: F(x) \neq F^*(x), \text{ for at least one } x$$

The test is conducted as follows. Let x_1, x_2, \dots, x_n represent a random sample of n independent observations of the random variable X . These n observations are grouped into c predetermined, mutually exclusive classes, where O_j denotes the number of observations in class j ($j = 1, 2, \dots, c$). Let

$$E_j = p_j^* n \quad (j = 1, 2, \dots, c) \quad (9.34)$$

where $E_j =$ the expected number of observations in class j when H_0 is true; that is,

$$p_j^* = \text{the probability of an observation } x \text{ being in class } j \text{ when } H_0 \text{ is true}$$

The test statistic for the chi-square goodness-of-fit test is given by

$$\chi^* = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} \quad (9.35)$$

An equivalent expression which is more convenient for use with a hand calculator is

$$\chi^* = \sum_{j=1}^c \frac{O_j^2}{E_j} - n \quad (9.36)$$

The exact distribution of χ^* is difficult to use; however, it can be approximated by a chi-square distribution with $(c - 1)$ degrees of freedom (the

†If $F^*(x)$ is not completely specified, see the discussion at the end of this section.

approximation improves as the sample size n increases). Thus, H_0 in Equation 9.32 is rejected if $\chi^* > \chi^2_{1-\alpha}(c-1)$, obtained from Table A4.

The chi-square goodness-of-fit test, as presented thus far, is somewhat arbitrary because no definite criterion for grouping the random sample into classes is specified. It is likely that different groupings will yield different values of the test statistic and possibly different conclusions, e.g., rejecting or not rejecting H_0 . Thus, before an example is presented, some guidelines for grouping observations into classes are given. Cochran (1952) suggests that none of the E_j should be less than 1 and no more than 20% of the E_j should be smaller than 5. However, more recent studies by Yarnold (1970) and Slakter (1973) suggest that Cochran's guidelines may be more restrictive than necessary. Yarnold (1970) states that if the number of classes c is 3 or more, and if r denotes the number of expectations less than 5, then the minimum expectation may be as small as $5r/c$. Slakter (1973) proposes that the number of classes can exceed the number of observations, which implies that the average expected value can be less than 1.0. In this text, we recommend that the observations be grouped in such a way that Yarnold's (1970) criterion is satisfied. This may require combining some groups if the E_j are small.

Example 9.16 A fuel fabrication facility routinely monitors its gaseous effluents by using an α -counter to measure the uranium content of air filters. The α -counter output is the number of counts observed in 1 min, which can be translated into a quantity of uranium given the appropriate calibration relationship. The facility purchased a new α -counter and must test the manufacturer's claim that the counter's random measurement errors follow a normal distribution with a standard deviation of 10 counts/min for samples yielding between 200 and 400 counts/min. A standard rated at 310 counts/min was measured 30 times, yielding the following counts:

317, 316, 314, 316, 278, 313, 300, 317, 287, 325,
 267, 299, 321, 283, 315, 280, 306, 354, 319, 342,
 270, 351, 314, 284, 316, 331, 296, 304, 300, 326

Use the chi-square goodness-of-fit test to test the manufacturer's claim at the $\alpha = 0.10$ level of significance.

One possible way to group the observations is by choosing the class boundaries to be the quartiles (Section 1.3.4) of the normal distribution with mean 310 and variance 100, i.e., 10^2 . By applying the methods described in Section 2.5.2, these are defined as follows:

$$\begin{aligned}x_{0.25} &= 310 + 10(z_{0.25}) = 310 + 10(-0.67) = 303.3 \\x_{0.50} &= 310 + 10(z_{0.50}) = 310 + 10(0) = 310.0 \\x_{0.75} &= 310 + 10(z_{0.75}) = 310 + 10(0.67) = 316.7\end{aligned}$$

where the values $z_{0.50} = 0.00$ and $z_{0.75} = 0.67$ are obtained from Table A3, and $z_{0.25} = -z_{0.75} = -0.67$. The observations are then grouped as shown below.

| Class | Range | Observed frequency (O_j) | Expected frequency (E_j) |
|-------|--------------------|------------------------------|------------------------------|
| 1 | $(-\infty, 303.3)$ | 11 | 7.5 |
| 2 | $(303.3, 310.0)$ | 2 | 7.5 |
| 3 | $(310.0, 316.7)$ | 7 | 7.5 |
| 4 | $(316.7, \infty)$ | 10 | 7.5 |

The test statistic is computed from Equation 9.35, which yields

$$\chi^* = \frac{(11 - 7.5)^2}{7.5} + \frac{(2 - 7.5)^2}{7.5} + \frac{(7 - 7.5)^2}{7.5} + \frac{(10 - 7.5)^2}{7.5} = 6.53$$

With $\alpha = 0.10$ and $c - 1 = 3$, the value $\chi_{0.90}^2(3) = 6.25$ is obtained from Table A4. Because $6.53 > 6.25$, the manufacturer's claim is rejected.

Although the null hypothesis that the counts are from a normal distribution with mean 310 and standard deviation 10 is rejected, it is possible that the data are from a normal distribution with a different mean and/or standard deviation. This possibility is considered in Example 9.17 (presented after the following discussion).

In the above presentation of the chi-square test, the reference distribution $F^*(x)$ is completely specified. Conover (1980) discusses modifications of the chi-square test when $F^*(x)$ is unspecified, except for the values of k parameters. These modifications require first grouping the observations and then estimating the k unspecified parameters by applying the method of moments to the grouped data. The test then proceeds as outlined previously, except that the distribution of χ^* is now approximated by a chi-square distribution with $c - k - 1$ degrees of freedom. This modified test is illustrated in the following example.

Example 9.17 Use the data from Example 9.16 to test the hypothesis that the measurements are from a normal distribution with unspecified

parameters. Use the modified chi-square test with the $\alpha = 0.10$ level of significance.

Because no parameter values are specified, the data are arbitrarily grouped into four classes as follows:

| Class | Class boundaries | Class midpoint, m_i | Class frequency, O_j |
|-------|----------------------|-----------------------|------------------------|
| 1 | $260 \leq x_i < 285$ | 272.5 | 6 |
| 2 | $285 \leq x_i < 310$ | 297.5 | 7 |
| 3 | $310 \leq x_i < 335$ | 322.5 | 14 |
| 4 | $335 \leq x_i < 360$ | 347.5 | 3 |

From Section 3.3.1, the method of moments estimates of the mean and variance are computed as

$$\begin{aligned}\hat{\mu} &= \sum_{j=1}^4 O_j m_j / n \\ &= [6(272.5) + 7(297.5) + 14(322.5) + 3(347.5)] / 30 \\ &= 309.167\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2 &= \left(\sum_{j=1}^4 O_j m_j^2 / n \right) - \hat{\mu}^2 \\ &= [6(272.5)^2 + 7(297.5)^2 \\ &\quad + 14(322.5)^2 + 3(347.5)^2] / 30 - (309.167)^2 \\ &= 96114.588 - (309.167)^2 = 530.35\end{aligned}$$

The expected values E_j are then computed for a normal distribution with mean 309.167 and variance 530.35 (i.e., standard deviation 23.03). This is done for the four classes defined above and for the "tails" of the distribution (classes 0 and 5). To illustrate the computations, consider class 1. First, by applying the methods presented in Section 2.5.2 with values obtained from Table A3, p_1 is computed as

$$\begin{aligned}
 p_1 &= \Pr(260 \leq X < 285) \\
 &= \Pr\left(\frac{260 - 309.167}{23.03} \leq Z < \frac{285 - 309.167}{23.03}\right) \\
 &= \Pr(-2.13 \leq Z < -1.05) \\
 &= \Pr(Z < -1.05) - \Pr(Z < -2.13) \\
 &= 0.14686 - 0.01659 = 0.13027
 \end{aligned}$$

Then

$$E_1 = np_1 = 30(0.13027) = 3.91$$

The other E_j values are similarly computed, and the results are displayed below.

| Class | Boundaries | p_j | E_j |
|-------|----------------------|---------|-------|
| 0 | $x_i < 260$ | 0.01659 | 0.50 |
| 1 | $260 \leq x_i < 285$ | 0.13027 | 3.91 |
| 2 | $285 \leq x_i < 310$ | 0.36909 | 11.07 |
| 3 | $310 \leq x_i < 335$ | 0.35269 | 10.58 |
| 4 | $335 \leq x_i < 360$ | 0.11781 | 3.53 |
| 5 | $x_i \geq 360$ | 0.01355 | 0.41 |

Because the values of E_j for classes 0 and 5 are quite small, they are combined with classes 1 and 4, respectively. The resulting values are tabulated below.

| Class | O_j | E_j |
|----------------------|-------|-------|
| $x_i < 285$ | 6 | 4.41 |
| $285 \leq x_i < 310$ | 7 | 11.07 |
| $310 \leq x_i < 335$ | 14 | 10.58 |
| $x_i \geq 335$ | 3 | 3.94 |

The test statistic is computed from Equation 9.36, which gives

$$\chi^* = \frac{6^2}{4.41} + \frac{7^2}{11.07} + \frac{14^2}{10.58} + \frac{3^2}{3.94} - 30 = 3.40$$

With $\alpha = 0.10$ and $c - k - 1 = 4 - 2 - 1 = 1$, the value $\chi_{0.90}^2(1) = 2.71$ is obtained from Table A4. Because $3.40 > 2.71$, reject the hypothesis that the measurements are from a normal distribution.

Note that in Example 9.16, application of the chi-square test led to rejection of the hypothesis that the measurements are from a normal distribution with mean 310 and standard deviation 10. But this did not rule out the possibility that the measurements are from a normal distribution with some other parameter values. In the present example, however, the modified chi-square test led to rejection of the hypothesis that the measurements are from a normal distribution with unspecified parameters.

9.6.2 Kolmogorov Test

The Kolmogorov test, also known as the Kolmogorov-Smirnov test, can be used to test goodness-of-fit for both discrete and continuous distributions. Let $F(x)$ be the true cdf of the random variable X , and let $F^*(x)$ be a completely specified cdf which is the hypothesized distribution. It is of interest to test the hypothesis given in Equation 9.33. The Kolmogorov test compares the **empirical cumulative distribution function** $G(x)$ from the random sample (Section 3.1.1) with the hypothesized cdf $F^*(x)$. The empirical cdf is a function of x , which equals the fraction of the observations x_i that are less than or equal to x for each x , $-\infty < x < \infty$. If there is not good agreement between $G(x)$ and $F^*(x)$, then H_0 in Equation 9.33 is rejected.

The Kolmogorov test is conducted as follows. Let the test statistic T^* be the greatest (denoted by "sup" for supremum) absolute distance between $G(x)$ and $F^*(x)$; that is,

$$T^* = \sup_x |F^*(x) - G(x)| \quad (9.37)$$

The supremum requires comparing $F^*(x)$ to $G(x)$ both just before and just after each step in $G(x)$. This is accomplished by computing both $|F^*(x_i) - G(x_i)|$ and $|F^*(x_i) - G(x_{i-1})|$ and then letting T^* be the largest of these absolute differences over all $i = 1, 2, \dots, n$. For a completely specified distribution, values of $F^*(x_i)$ are computed by applying Equation 2.3 or 2.5. For a normal distribution $F^*(x)$ is defined by Equations 2.33

and 2.35, with computations illustrated in Example 2.5. H_0 is rejected at the α level of significance if T^* exceeds $w_{1-\alpha}$, obtained from Table A17.

Example 9.18 Rework the problem outlined in Example 9.16 by using the Kolmogorov test instead of the chi-square test. Use the $\alpha = 0.10$ level of significance.

The empirical cdf and values of the cdf $F^*(x)$ for the hypothesized normal distribution $N(310, (10)^2)$ are displayed in Table 9.14. From Equation 9.37, the test statistic, $T^* = 0.222$, is the largest value from the last two columns of Table 9.14. With $n = 30$ and $\alpha = 0.10$, the value $w_{0.90} =$

TABLE 9.14
Empirical and Hypothesized cdf Values

| Observation number | Ordered values of X (x_i) | $G(x_i)$ | $F^*(x_i)$ | $ F^*(x_i) - G(x_i) $ | $ F^*(x_i) - G(x_{i-1}) $ |
|--------------------|---------------------------------|----------|------------|-----------------------|---------------------------|
| 1 | 267 | 1/30 | 0.000 | 0.033 | 0 |
| 2 | 270 | 2/30 | 0.000 | 0.067 | 0.033 |
| 3 | 278 | 3/30 | 0.000 | 0.100 | 0.067 |
| 4 | 280 | 4/30 | 0.001 | 0.132 | 0.099 |
| 5 | 283 | 5/30 | 0.003 | 0.164 | 0.130 |
| 6 | 284 | 6/30 | 0.005 | 0.195 | 0.162 |
| 7 | 287 | 7/30 | 0.011 | 0.222 | 0.189 |
| 8 | 296 | 8/30 | 0.081 | 0.186 | 0.152 |
| 9 | 299 | 9/30 | 0.136 | 0.164 | 0.131 |
| 10 | 300 | — | — | — | — |
| 11 | 300 | 11/30 | 0.159 | 0.208 | 0.141 |
| 12 | 304 | 12/30 | 0.274 | 0.126 | 0.093 |
| 13 | 306 | 13/30 | 0.345 | 0.088 | 0.055 |
| 14 | 313 | 14/30 | 0.618 | 0.151 | 0.185 |
| 15 | 314 | — | — | — | — |
| 16 | 314 | 16/30 | 0.655 | 0.122 | 0.188 |
| 17 | 315 | 17/30 | 0.691 | 0.124 | 0.158 |
| 18 | 316 | — | — | — | — |
| 19 | 316 | — | — | — | — |
| 20 | 316 | 20/30 | 0.726 | 0.059 | 0.159 |
| 21 | 317 | — | — | — | — |
| 22 | 317 | 22/30 | 0.758 | 0.025 | 0.091 |
| 23 | 319 | 23/30 | 0.816 | 0.049 | 0.083 |
| 24 | 321 | 24/30 | 0.864 | 0.064 | 0.097 |
| 25 | 325 | 25/30 | 0.933 | 0.100 | 0.133 |
| 26 | 326 | 26/30 | 0.945 | 0.078 | 0.112 |
| 27 | 331 | 27/30 | 0.982 | 0.082 | 0.115 |
| 28 | 342 | 28/30 | 1.000 | 0.067 | 0.100 |
| 29 | 351 | 29/30 | 1.000 | 0.033 | 0.067 |
| 30 | 354 | 1 | 1.000 | 0.000 | 0.033 |

0.218 is obtained from Table A17. Because $0.222 > 0.218$, $H_0: F(x) = F^*(x)$ is rejected at the $\alpha = 0.10$ level of significance. This indicates that the measurements are probably not from the normal distribution with mean 310 and standard deviation 10. This is consistent with the conclusion reached in Example 9.16.

Examples 9.16 and 9.18 are based on the same data to allow comparison of the results. Both the chi-square test and the Kolmogorov test result in rejecting H_0 at the $\alpha = 0.10$ level of significance. In general, Conover (1980) suggests that the Kolmogorov test may be preferred over the chi-square test if the sample size is small, because exact critical values are readily available for the Kolmogorov test. Conover also states that the Kolmogorov test is more powerful than the chi-square test for many situations. For further details and comparisons, see Slakter (1965).

9.6.3 Tests for Normality

The introduction to this section mentions a class of goodness-of-fit tests that do not require the hypothesized distribution $F^*(x)$ to be completely specified. Two tests of this type are presented for testing the hypothesis that the data are from a normal distribution with unspecified parameters. These are the **Lilliefors test**, and the **Shapiro-Wilk test**. For each of these tests, $F^*(x)$ is a normal distribution, where μ and σ^2 are not specified. The observations x_i are used to estimate μ and σ^2 .

9.6.3.1 The Lilliefors Test for Normality

The Lilliefors test for normality, developed by Lilliefors (1967), is a modification of the Kolmogorov test. It is of interest to test the hypothesis

H_0 : The random sample comes from a normal distribution with unspecified mean and variance

against the alternative

(9.38)

H_A : The distribution is not normal

The test is conducted as follows. Let x_1, x_2, \dots, x_n represent the values of a random sample of n independent observations of the random variable X with unknown cdf $F(x)$. Compute the standardized sample values

$$z_i = \frac{x_i - \bar{x}}{s} \quad (i = 1, 2, \dots, n)$$

where \bar{x} and s are the sample mean and sample standard deviation, respectively. The test consists of letting $F^*(z)$ be the standard normal cdf and then comparing it to the empirical cdf of the z_i 's, denoted by $G(z)$. The Lilliefors test statistic is the greatest difference between $F^*(z)$ and $G(z)$; that is,

$$T^* = \sup_z |F^*(z) - G(z)| \quad (9.39)$$

T^* is the largest of all values $|F^*(z_i) - G(z_i)|$ or $|F^*(z_i) - G(z_{i-1})|$. Values of $F^*(z_i)$ are obtained from Table A3. H_0 is rejected at the α level of significance if T^* exceeds $w_{1-\alpha}$, obtained from Table A18.

Example 9.19† Percent uranium values for 17 cans of ammonium diuranate (ADU) scrap are as follows:

| | | |
|------|------|------|
| 35.5 | 78.2 | 29.4 |
| 79.4 | 37.1 | 29.8 |
| 35.2 | 48.4 | 28.4 |
| 40.1 | 28.6 | 23.4 |
| 25.0 | 75.5 | 77.0 |
| 78.5 | 34.3 | |

Use the Lilliefors test to test the hypothesis that these observations are from a normal distribution. Use the $\alpha = 0.05$ level of significance.

For the 17 observations, the sample mean and standard deviation are $\bar{x} = 46.11$ and $s = 21.83$. The ordered x_i and the corresponding values of z_i , $F^*(z_i)$, $G(z_i)$, $|F^*(z_i) - G(z_i)|$ and $|F^*(z_i) - G(z_{i-1})|$ are displayed in Table 9.15.

From Table 9.15, the Lilliefors test statistic is $T^* = 0.257$. For $n = 17$ and $\alpha = 0.05$, the value $w_{0.95} = 0.206$ is obtained from Table A18. Because $0.257 > 0.206$, the null hypothesis of normality is rejected at the $\alpha = 0.05$ significance level. This indicates that the observations are probably not from a normal distribution.

9.6.3.2 The Shapiro-Wilk Test for Normality

The **Shapiro-Wilk test** for normality was developed by Shapiro and Wilk (1965, 1968). The test requires a table of coefficients to compute the test statistic in addition to a table of critical values.

†The data for this example are from Jaech (1973).

TABLE 9.15
Lilliefors Test Computations

| Ordered
x_i | z_i | $F^*(z_i)$ | $G(z_i)$ | $ F^*(z_i) - G(z_i) $ | $ F^*(z_i) - G(z_{i-1}) $ |
|------------------|-------|------------|----------|-----------------------|---------------------------|
| 23.4 | -1.04 | 0.149 | 1/17 | 0.090 | 0.149 |
| 25.0 | -0.97 | 0.166 | 2/17 | 0.048 | 0.107 |
| 28.4 | -0.81 | 0.209 | 3/17 | 0.033 | 0.091 |
| 28.6 | -0.80 | 0.212 | 4/17 | 0.023 | 0.036 |
| 29.4 | -0.77 | 0.221 | 5/17 | 0.073 | 0.014 |
| 29.8 | -0.75 | 0.227 | 6/17 | 0.126 | 0.067 |
| 34.3 | -0.54 | 0.295 | 7/17 | 0.117 | 0.058 |
| 35.2 | -0.50 | 0.309 | 8/17 | 0.162 | 0.103 |
| 35.5 | -0.49 | 0.312 | 9/17 | 0.217 | 0.159 |
| 37.1 | -0.41 | 0.341 | 10/17 | 0.247 | 0.188 |
| 40.1 | -0.28 | 0.390 | 11/17 | 0.257 | 0.198 |
| 48.4 | 0.11 | 0.544 | 12/17 | 0.162 | 0.103 |
| 75.5 | 1.35 | 0.911 | 13/17 | 0.146 | 0.205 |
| 77.0 | 1.42 | 0.922 | 14/17 | 0.098 | 0.157 |
| 78.2 | 1.47 | 0.929 | 15/17 | 0.047 | 0.105 |
| 78.5 | 1.48 | 0.931 | 16/17 | 0.010 | 0.049 |
| 79.4 | 1.53 | 0.937 | 1 | 0.063 | 0.004 |

The Shapiro-Wilk test of the hypothesis given in Equation 9.38 is conducted as follows. Let x_1, x_2, \dots, x_n represent the values of a random sample of n independent observations on the random variable X . Order the sample observations from smallest to largest

$$x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(n)}$$

where $x^{(i)}$ denotes the i^{th} order statistic (i.e., the i^{th} ordered observation). Then the test statistic is given by

$$W^* = \frac{b^2}{(n-1)s^2} \quad (9.40)$$

where

$$b = \sum_{i=1}^k a_i (x^{(n-i+1)} - x^{(i)})$$

s = the sample standard deviation

k = $n/2$ if n is even or $(n-1)/2$ if n is odd

a_i = coefficients (which depend on n) obtained from Table A19

The null hypothesis of normality is rejected at the α level of significance if W^* is less than the critical value w_α obtained from Table A20. This table

includes critical values for $n \leq 50$. Critical values for larger values of n are not available. One of the other tests for normality is recommended when the sample size is larger than 50.

In addition to the original articles by Shapiro and Wilk, detailed descriptions of the Shapiro-Wilk test are given by Conover (1980), Hahn and Shapiro (1968), and Jaech (1973). The reader is referred to these sources for additional examples and discussion.

Example 9.20 Rework the problem outlined in Example 9.19 by using the Shapiro-Wilk test.

The ordered data and other computational steps are summarized in Table 9.16. The sample standard deviation is 21.83. The values of a_i are obtained from Table A19 for $n = 17$.

TABLE 9.16
Calculations for Shapiro-Wilk Test for Normality

| i | Ordered data, $x^{(i)}$ | $x^{(n-i+1)}$ | $x^{(n-i+1)} - x^{(i)}$ | a_i | $a_i x^{(n-i+1)} - x^{(i)}$ |
|-----|-------------------------|---------------|-------------------------|--------|-----------------------------|
| 1 | 23.4 | 79.4 | 56.0 | 0.4968 | 27.8208 |
| 2 | 25.0 | 78.5 | 53.5 | 0.3273 | 17.5106 |
| 3 | 28.4 | 78.2 | 49.8 | 0.2540 | 12.6492 |
| 4 | 28.6 | 77.0 | 48.4 | 0.1988 | 9.6219 |
| 5 | 29.4 | 75.5 | 46.1 | 0.1524 | 7.0256 |
| 6 | 29.8 | 48.4 | 18.6 | 0.1109 | 2.0627 |
| 7 | 34.3 | 40.1 | 5.8 | 0.0725 | 0.4205 |
| 8 | 35.2 | 37.1 | 1.9 | 0.0359 | 0.0682 |
| 9 | 35.5 | | | | |
| 10 | 37.1 | | | | $b = 77.1796$ |
| 11 | 40.1 | | | | |
| 12 | 48.4 | | | | |
| 13 | 75.5 | | | | |
| 14 | 77.0 | | | | |
| 15 | 78.2 | | | | |
| 16 | 78.5 | | | | |
| 17 | 79.4 | | | | |

The test statistic is computed from Equation 9.40, which gives

$$W^* = \frac{(77.1796)^2}{(17 - 1)(21.83)^2} = 0.781$$

With $n = 17$ and $\alpha = 0.05$, the value $w_{0.05} = 0.892$ is obtained from Table A20. Because $0.781 < 0.892$, the null hypothesis is rejected. This indicates that the observations are probably not from a normal distribution. This agrees with the conclusion reached in Example 9.19.

In this presentation of goodness-of-fit tests, three tests have been applied to test the hypothesis that the observations in a random sample are from an unspecified normal distribution. These are the modified chi-square test, the Lilliefors test and the Shapiro-Wilk test. The question arises as to which of these tests to use in practice.

Based on power comparison studies by Shapiro, Wilk, and Chen (1968) and LeBrecque (1977), the Shapiro-Wilk test appears to be superior to the chi-square and Lilliefors tests in many situations. Although there are differences in power among these tests, in practice these differences are probably of little concern. The reason goodness-of-fit tests for normality are performed is usually to determine whether the assumptions underlying parametric statistical analysis techniques are valid. If a goodness-of-fit test casts doubt on the validity of the normality assumption, use of non-parametric techniques is recommended.

It is important to note that in practice, observations probably never come from a distribution that can be exactly characterized by a particular (hypothesized) distribution. As a consequence, almost any goodness-of-fit test will result in rejection of the null hypothesis if the number of observations is very large. Thus, the outcome of such a test should be carefully interpreted. Goodness-of-fit tests provide a criterion for determining whether the agreement between the true and hypothesized distributions is close enough that the hypothesized distribution provides a satisfactory approximation to the true distribution. If the approximation is deemed satisfactory, then statistical methods based on the hypothesized distribution can be applied with some assurance that the results (inferences) are valid.

REFERENCES

- Bradley, J. V., 1968, *Distribution-Free Statistical Tests*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- Cochran, W. G., 1952, The χ^2 Test of Goodness of Fit, *Ann. Math. Stat.*, **23**: 315-345.
- Conover, W. J., 1971, *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York.
- , 1980, *Practical Nonparametric Statistics*, 2nd Ed., John Wiley & Sons, Inc., New York.
- , and R. L. Iman, 1978, Some Exact Tables for the Squared Ranks Test, *Commun. Stat., Pt. B: Simul. Comput.*, **B7**: 491-513.
- Crouse, C. F., 1968, A Distribution-Free Method of Analyzing a 2^m Factorial Experiment, *South African Statistics Journal*, **2**: 101-108.
- Draper, N. R., and H. Smith, 1981, *Applied Regression Analysis*, 2nd Edition, John Wiley & Sons, Inc., New York.

- Hahn, G. J., and S. S. Shapiro, 1968, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York.
- Hora, S. C., and W. J. Conover, 1981, *The Limiting Null Distribution of the F-Statistic in Factorial Designs with Rank-Score Transformed Data*, Texas Tech University Statistics Department Technical Report, Lubbock, Texas.
- Iman, R. L., S. C. Hora, and W. J. Conover, 1981, *A Comparison of Asymptotically Distribution Free Procedures for the Analysis of Complete Blocks*, Texas Tech University Statistics Department Technical Report, Lubbock, Texas.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- LaBrecque, J., 1977, Goodness-of-Fit Tests Based on Nonlinearity in Probability Plots, *Technometrics*, **19**: 293-306.
- Lehmann, E. L., 1975, *Nonparametrics: Statistical Methods Based on Ranks*, Holden Day, San Francisco.
- Lilliefors, H. W., 1967, On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *J. Am. Stat. Assoc.*, **62**: 399-402.
- McDonald, P., 1971, The Analysis of a 2nd Experiment by Means of Ranks, *J. R. Stat. Soc., Ser. C*, **20**: 259-275.
- Shapiro, S. S., and M. B. Wilk, 1965, An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, **52**: 591-611.
- , and M. B. Wilk, 1968, Approximations for the Null Distribution of the W Statistic, *Technometrics*, **10**: 861-866.
- , M. B. Wilk, and H. J. Chen, 1968, A Comparative Study of Various Tests for Normality, *J. Am. Stat. Assoc.*, **63**: 1343-1372.
- Slakter, M. J., 1965, A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Validity, *J. Am. Stat. Assoc.*, **60**: 854-858.
- , 1973, Large Values for the Number of Groups with the Pearson Chi-Squared Goodness-of-Fit Test, *Biometrika*, **60**: 420-421.
- Swed, F. S., and C. Eisenhart, 1943, Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *Ann. Math. Stat.*, **14**: 66-87.
- Yarnold, J. K., 1970, The Minimum Expectations in χ^2 Goodness-to-Fit Tests and the Accuracy of Approximations for the Null Distribution, *J. Am. Stat. Assoc.*, **65**: 864-886.

CHAPTER 10

Simultaneous Inference and Detection of Outliers

10.0 INTRODUCTION

In this chapter, the statistical concept introduced in Chapter 5 with the Duncan, Bonferroni, or Scheffé procedures is further expanded. This is the concept of simultaneous inference, also called multiple comparisons. In the analysis of variance when the null hypothesis of no differences among group means is rejected, possible alternative hypotheses are often examined so that those best explaining the data can be selected. However, using data to suggest hypotheses and/or using the same data many times to test hypotheses violates the required assumptions of independent data sets and *a priori* hypotheses. The question that immediately arises is whether the same results would be obtained if the experiment were repeated. By randomness alone, many data sets will have some anomalies that appear to be significant. The techniques of simultaneous inference protect against finding, by chance, a spurious result due to a particular pattern of randomness that is manifest in only one data set. This chapter presents some of the statistical concepts that form the foundation of simultaneous inference.

This chapter also presents statistical methods that are used to test for outliers in a set of data. An **outlier** is an observed or measured value that is unusually large or small compared to other values obtained under the same conditions, or when compared with prior knowledge of the range of values that would be expected under similar conditions. Statistical methods for detecting outliers involve simultaneous inference because the need for such methods is determined after examining the data set. Although there is overlap in the underlying philosophy and theory of simultaneous inference and testing for outliers (e.g., the same set of data might be used repeatedly to test various hypotheses suggested by the data), the subject of testing for outliers is broad enough and sufficiently important to be treated separately.

As with many of the other topics presented in this book, a large body of published literature exists for both simultaneous inference and testing for outliers. Thus, this chapter is intended to be a brief introduction to the topics. The reader interested in studying either of these topics in greater detail should consult some of the listed references.

10.1 THE SIMULTANEOUS INFERENCE PROBLEM

Often, in practice, a single set of data is used to investigate many issues. For example, suppose that data are available on the percent ^{235}U in 20 batches of uranium oxide. A production supervisor is interested in comparing individual batches with specifications, making comparisons among batches, and comparing groups of batches that might be blended. An inspector, on the other hand, is interested in comparing results of verification measurements with the operator's values for the 20 batches. Thus, not only will multiple statistical comparisons be performed, but the particular comparisons to be made will be chosen after the data are reviewed. Although such practices violate the assumptions underlying statistical testing, it is understandable that the production supervisor and the inspector are interested in extracting as much information as possible from these data. On the other hand, if a multiplicity of tests is applied to the same data set, the chances increase of finding at least one statistically significant difference even if no true differences exist. Before taking action on the basis of results from one of a multiplicity of tests, allowance for the possible errors in using a multiplicity of tests must be made.

Suppose that verification data are available on 20 batches of uranium oxide. Each batch is tested at the 0.05 significance level to check whether the percent ^{235}U agrees with the operator's values. The difference between the two estimates for one of the batches is found to be significantly different from zero at the 0.05 significance level. The 0.05 significance level implies that, in the long run, a Type I error will be made 5% of the time in the inference from a test. If 20 such tests are performed, it is expected that one significant difference would be found when no true differences exist. However, the probability that one or more Type I errors will be made in such a situation is $1 - (0.95)^{20} = 0.64$ (Section 10.2.1). In this example, the correct approach is to simultaneously make the 20 comparisons in such a way that the simultaneous inference can be made with an overall Type I error probability of 0.05. The procedures of simultaneous inference provide statistical tools for making multiple tests at a specified overall significance level. In the example above, rather than performing 20 tests each at the 0.05 significance level, one test of 20 comparisons would be performed with an overall 0.05 significance level.

10.1.1 Error Rates

When a number of statistical statements are considered simultaneously, they are collectively called a **family** of statements. The statements can be significance statements from hypothesis testing, statements about confidence intervals, or a combination of these. Let N be the number of statements in the family, and let n be the number of incorrect statements in the family. The **error rate** of the family is n/N . The error rate is a random

variable whose distribution depends upon the test statistics used and the kinds of statistical statements in the family. To use the concept of error rate to assess the collective statistical significance of statements in the family, a summary criterion is needed that does not depend upon the distribution of the error rate. Two such criteria are (1) the probability of a nonzero error rate, and (2) the expected error rate. In simultaneous inference, probabilities associated with one or both of these criteria are fixed, and then individual statements within the family are adjusted so that the probabilities are satisfied. More details on this approach to simultaneous inference are given in Miller (1966).

The first criterion, the probability of a nonzero error rate, is defined equivalently as the probability that n/N or n is greater than zero. To have a specified probability of a nonzero error rate for a large family, the probability of an error for each statement in the family must be much smaller than that specified for the family. The simultaneous significance level is equal to the probability of the nonzero error rate.

The **Bonferroni inequality** gives a bound on how much the individual statement probabilities have to be reduced to yield a specified probability of a nonzero error rate. The Bonferroni inequality is

$$\Pr(S_1 S_2 \dots S_N) \geq 1 - [\Pr(S'_1) + \Pr(S'_2) + \dots + \Pr(S'_N)] \quad (10.1)$$

where S_i is a statement in a family of size N , and S'_i is its complement. This inequality can be used to determine how much to adjust probabilities corresponding to individual statements to achieve an upper bound on the simultaneous (collective or overall) significance level.

Example 10.1 Suppose that there are three material control areas within a nuclear fuel facility. Each control area has been tested separately for material balance. From this evaluation, the probability is 0.98 that no material is missing from Area 1, 0.96 for Area 2, and 0.975 for Area 3. What is the plant-wide probability that no loss has occurred?

Let S_i represent the statement that no loss has occurred from Area i . Then S'_i is its complement, the statement that a loss has occurred. Because $\Pr(S_1) = 0.98$, it follows that $\Pr(S'_1) = 0.02$. Likewise, $\Pr(S'_2) = 0.04$ and $\Pr(S'_3) = 0.025$. $\Pr(S_1 S_2 S_3)$ is the simultaneous probability that no loss has occurred from any of the three material control areas. Applying the Bonferroni inequality, given by Equation 10.1,

$$\Pr(S_1 S_2 S_3) \geq 1.0 - (0.02 + 0.04 + 0.025) = 0.915$$

Thus, the Bonferroni inequality indicates that the probability is at least 0.915 that there is no plant-wide loss.

This example illustrates how the Bonferroni inequality provides a bound on the probability for a simultaneous statement. There are other inequalities available that result in tighter bounds than the Bonferroni, for example that of Gabriel (1978). These are often more difficult to implement, and the Bonferroni inequality has been found to be close to the true value if the $\Pr(S'_i)$ are small and N is not too large. In many safeguards applications, however, N may be quite large, making the Bonferroni inequality too conservative.

The second criterion for assessing the overall error rate in a family is the expected error rate. The expected error rate has the desirable property of being the average of the marginal error rates. The expected error rate for Example 10.1 is the average of 0.02, 0.04 and 0.025; that is, 0.028. This average is interpreted differently than the 0.915 bound obtained from the Bonferroni inequality. One minus the expected error rate, $(1 - 0.028) = 0.972$, is a measure of average performance over the three material control areas; the Bonferroni value 0.915 is a lower bound on the probability that no loss has occurred in any of the three control areas.

10.1.2 Allocation of Error

The two criteria presented in the previous section for assessing simultaneous probabilities give methods for combining probabilities of individual statements within a family into a single probability for the entire family. This process can be used in reverse; that is, a probability for the family can be specified and the criteria can be used to determine bounds on probabilities of individual statements. This reverse process is typically used for comparing group means in analysis of variance applications (Section 5.1.4). In an analysis of variance, the family is typically large, such as the number of possible ways of comparing group means. Rather than making all possible comparisons, this problem is typically solved in the following way: (1) a subset of the family of probability statements is chosen (such as those statements that include only pairwise comparisons of means), and (2) the probability is allocated equally among the statements in the subset.

Example 10.2 Consider a situation similar to the one described in Example 10.1; that is, a nuclear fuel fabrication facility has three material control areas. Assume that inventory balances are performed by a single

team which inventories one control area in each of three successive months. It is necessary to report a plant-wide possible loss as soon as it is detected rather than to wait until all three inventories are completed. A protocol is needed that results in a plant-wide false alarm rate (significance level) of 0.10. Using equal allocation of error and the Bonferroni inequality, the protocol calls for an alarm whenever any single inventory shows a discrepancy at the 0.033 significance level, calculated as follows: Let S_i represent the event of no alarm in Area i when in fact, no material is missing. The plant-wide probability of no alarm when no material is missing is $\Pr(S_1 S_2 S_3) = 0.90$. The required false alarm rate can be assured by setting the right-hand side of Equation 10.1 equal to 0.90. For equal allocation, $\Pr(S_1) = \Pr(S_2) = \Pr(S_3)$. By using Equation 10.1, $0.90 = 1.0 - [\Pr(S_1) + \Pr(S_2) + \Pr(S_3)] = 1.0 - 3\Pr(S_i)$. Solving for $\Pr(S_i)$ gives $\Pr(S_i) = (1.0 - 0.9)/3 = 0.033$ for $i = 1, 2, 3$.

The equal allocation concept illustrated in Example 10.2 can be generalized. Let $1/4$ be the simultaneous (family) significance level desired and assume that n statistical tests are to be made within the family. Then if the significance level of each of the tests is at most $(1/4)/n$, the Bonferroni inequality is satisfied. This is the rule used in Section 5.1.4 for comparing group means.

There are often compelling reasons for not equally allocating the errors. In Example 10.2, suppose that one of the control areas contains only containers of fuel pellets and another only uranium oxide powder and scrap in containers. It is reasonable to assume that a given quantity of uranium pellets is easier to divert than uranium oxide powder or scrap. Thus, the protocol should allocate the error to provide greater protection against loss of pellets; that is, a larger significance level (resulting in a more sensitive test) should be used for this area.

The choice between expected error rate or nonzero error rate as a basis for allocating errors depends on the purpose of allocation and upon convention. Statisticians conventionally use a nonzero error rate for each data set analyzed, and use an expected (average) error rate for multiple data sets. As an example, a regional inspector who has responsibility for controlling SNM at several facilities might require using a nonzero error rate criterion for no diversion of material for each facility and an expected error rate criterion among facilities within the region.

10.2 SIMULTANEOUS INFERENCE METHODS

The family of statements that are considered simultaneously can be of two types: those containing correlated statements or those containing independent statements. When statements of a family are known to be

independent, exact simultaneous significance levels can be derived. If correlations or dependencies are present, they influence the significance level, and only bounds can be found for simultaneous significance levels.

Many situations in nuclear material control are composed of independent statements. There are two easily identifiable situations in which independence is evident. First, independence is assumed when groups of data are separated in time or space. For example, an inspector might wish to make a simultaneous statement about the inventory of fuel assemblies at two power plants in his region. The inventories of these two power plants are assumed to be independent. Second, independence is sometimes achieved by removing common factors from data sets. For example, if concentrations in evaporation ponds have annual cycles due to weather conditions, cyclic effects are removed by subtracting an appropriate function of time. Then measurements on evaporation ponds are compared by methods for independent data.

In addition, when a number of independent tests of significance are reviewed, it sometimes happens that none are individually significant, but the aggregate gives the impression that simultaneously the individual levels of significance are consistently lower or higher than would be obtained by chance alone. Some of the techniques presented in this section yield exact probability statements for independent aggregates. As mentioned previously, however, correlations will influence the significance levels. Therefore, if independence cannot be assumed, or if there is some doubt about independence, the methods that allow for correlations should be used.

10.2.1 Repeated Normal Statistics

When the data are assumed to be normally distributed, and when the standard deviation is known theoretically (or has been estimated to an accuracy that has led to the use of a value as a theoretical result) rather than computed from data, the normal distribution is used to construct confidence intervals for means or for a linear combination of means. Consider a family of N confidence interval statements with each confidence interval derived from a normal distribution. Let S_i represent the event that the i^{th} confidence statement is true, for $i = 1, 2, \dots, N$, and assume that the confidence intervals are constructed at the $100p_i\%$ confidence level. Then the nonzero error rate for the i^{th} statement is $1 - p_i$. If the S_i are not assumed to be independent, then a lower bound for $\Pr(S_1 S_2 \dots S_N)$ is given by Equation 10.1, which can be expressed in the form

$$\Pr(S_1 S_2 \dots S_N) \geq 1 - \sum_{i=1}^N (1 - p_i)$$

Thus, an upper bound for the probability of a nonzero error rate for the family is

$$P_{ER} = 1 - \Pr(S_1 S_2 \dots S_N) \leq \sum_{i=1}^N (1 - p_i) \quad (10.2)$$

However, if the S_i are independent, then

$$\Pr(S_1 S_2 \dots S_N) = p_1 p_2 \dots p_N$$

and the probability of a nonzero error rate for the family is

$$P_{ER} = 1 - \Pr(S_1 S_2 \dots S_N) = 1 - p_1 p_2 \dots p_N \quad (10.3)$$

If all p_i are equal, say $p_i = p$, Equation 10.3 becomes $1 - p^N$.

Example 10.3 Suppose that the quality control staff at a fuel fabrication plant keeps the scale used to weigh fuel pellets calibrated so a fuel pellet weight obtained from this scale has probability 0.95 of being within 0.1 g of the true weight of the pellet. What is the confidence level for the statement that all 10 pellets sampled from any tray of pellets will have their weights reported to within 0.1 g of their true weights?

If the intervals are independent and the scale readings are independent, the nonzero error rate (computed from Equation 10.3) is $1 - (0.95)^{10} = 0.40$, and the simultaneous statement can be made at the 60% confidence level.

If a new scale is purchased and a pellet weight read from the new scale has probability 0.995 of being within 0.1 g of the true pellet weight, then $\Pr(S_1 S_2 \dots S_{10}) = (0.995)^{10} = 0.951$, and the quality control staff can be 95% confident that all 10 sample pellet weights from any tray will be simultaneously accurate to within 0.1 g of their true weights.

10.2.2 Bonferroni t Statistics

Whenever the standard deviation is estimated from data being analyzed, Student's t-distribution is used instead of the normal to construct confidence intervals for means. In Chapter 3, it is shown that a $100(1-\alpha)\%$ confidence interval for the mean, based on a sample of size n , is constructed as $\bar{x} \pm t_{1-\alpha/2}(n-1)(s/\sqrt{n})$, where \bar{x} and s are the sample mean and sample standard deviation, and $t_{1-\alpha/2}(n-1)$ is obtained from Table A5.

The principle of equal allocation of errors (Section 10.1.2) can be used to make a simultaneous confidence statement for k samples. Let \bar{x}_i , s_i , and n_i be the mean, standard deviation, and sample size of the i^{th} sample, $i = 1, 2, \dots, k$. The sample results may be correlated. The formula

$$\bar{x}_i \pm t_{(1-\alpha/2k)}(n_i - 1)(s_i/\sqrt{n_i}) \quad (10.4)$$

gives a confidence interval for the i^{th} mean with a simultaneous confidence level greater than or equal to $100(1-\alpha)\%$ over the k samples. This procedure results in individual confidence intervals that are wider than those that would result if $t_{(1-\alpha/2)}(n_i - 1)$ had been used.

10.2.3 Studentized Maximum Modulus

A number of studentized statistics are found in the simultaneous inference literature. All methods based on studentized statistics require special tables, and many of the tables are not conveniently available. The studentized maximum modulus method, named for the distribution from which tabled values are derived, is presented here because a good table of modulus values has been published by Bechhofer and Dunnett (1982).

For the studentized maximum modulus method, all sample groups in the family of statements must be independent and must have the same standard deviation. Thus, this procedure differs from the repeated normal procedure that requires known standard deviations and from the Bonferroni t procedure that allows different standard deviations estimated from the data. The studentized maximum modulus gives shorter confidence intervals than the Bonferroni t intervals and is used when the variances can be assumed to be equal.

Consider N samples with means $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$ and sample sizes n_1, n_2, \dots, n_N , and with unknown underlying variance σ^2 . Let s^2 (the within-class mean square in analysis of variance terminology) be an independent estimate of σ^2 with ν degrees of freedom. A simultaneous set of N confidence intervals with a simultaneous confidence level of at least $100(1 - \alpha)\%$ is constructed from the formula

$$\bar{x}_i \pm m_{\alpha, N, \nu}(s/\sqrt{n_i}) \quad (10.5)$$

where $m_{\alpha, N, \nu}$ is found in a table of the studentized maximum modulus, such as that in Bechhofer and Dunnett (1982).

10.2.4 Chi-Square Sums

Two theorems from mathematical statistics allow the chi-square distribution to be applied to simultaneous inference problems where the state-

ments of a family are known to be independent. The first of these theorems and an example of its application are presented below.

Theorem 10.1 The sum of independent chi-square variables is also a chi-square variable, with degrees of freedom equal to the sum of the degrees of freedom of the individual variables in the sum.

Example 10.4 Suppose that a regional inspector is interested in the accuracy and precision of six scales used to weigh fuel pellets at facilities in his region. On each trip to each facility he weighs and records the weight of the same calibration standard.† The first step in his data analysis is to apply the chi-square goodness-of-fit test (Section 9.6.1) to determine whether the weighing errors for each scale are normally distributed. The six chi-square values (and associated degrees of freedom) are 6.8(4), 8.3(5), 7.8(5), 6.5(4), 10.1(6), and 7.2(4). Individually, each of the goodness-of-fit tests is significant at a level between 0.1 and 0.2 (Table A4); thus, at the 0.05 level of significance, no single scale shows a lack of normality in weighing errors. However, because all six of the chi-square values have similar levels of significance, these results should be further investigated. If the results were due to chance alone, it is expected that the significance levels would be more widely distributed between zero and unity. Summing the chi-square values gives 46.7 with 28 degrees of freedom, which is significant at the 0.025 level (Table A4). This indicates that the weighing errors for the scales are probably not all from normal distributions.

This example illustrates how sometimes a simultaneous test detects effects that are too subtle to be found by individual tests. Note, however, that these results do not indicate the cause of the overall lack of fit. The inspector will have to design a more sophisticated experiment to determine how the weighing error distributions differ from a normal distribution.

The second theorem allows extension of this chi-square sum technique.

Theorem 10.2 The square of a standardized normal variable is a chi-square distributed variable with one degree of freedom.

Many test statistics, such as those used in nonparametric tests (Chapter 9), are either asymptotically normally distributed or can be transformed to be approximately normally distributed. For example,

†Note that these data may not be rounded. See Jaech (1974).

Student's t -distribution with degrees of freedom greater than 30 is adequately approximated by a standard normal distribution. Whenever several test results are available and each is a standardized normal variable, Theorem 10.2 followed by Theorem 10.1 can be used to make a simultaneous probability statement about the pooled test results.

Example 10.5 In Example 9.2, assume that measurements of mean percentage UO_2 in lots of uranium oxide powder are tested for randomness using the sign test, and that sample sizes are large enough to use the normal approximation given in Equation 9.2. Further assume that five batches are tested, each having a different mean percentage UO_2 , so the data cannot be combined. The standard normal deviates from testing for too many runs are found to be 1.46, 1.51, 1.60, 1.39, and 1.54. Individually, these tests indicate nonrandomness at significance levels between 0.055 and 0.082 (Table A3). Thus, the individual hypotheses of randomness would not be rejected at the typically chosen 0.05 significance level. Note, however, that the test statistics and their significance levels are quite similar, and by chance alone, it is unusual to observe five independent test statistics with practically the same significance level. By Theorem 10.2, the square of each standard normal deviate has a chi-square distribution with one degree of freedom. Squaring and summing these standard normal deviates yields 11.28 which, by Theorem 10.1, is an observation from the chi-square distribution with 5 degrees of freedom. The value $\chi_{0.95}^2(5) = 11.1$ is obtained from Table A4. Because $11.28 > 11.1$, the five runs tests simultaneously indicate nonrandomness at the 0.05 level of significance.

10.2.5 Fisher's Method

Fisher (1970) showed that a function of the significance level associated with a statement has approximately a chi-square distribution with two degrees of freedom. This, in combination with Theorem 10.1, forms the basis of **Fisher's combination of probabilities test of significance**. Let p_i represent the probability (significance level) associated with the i^{th} statement in a family of n independent statements. Then the quantity

$$\chi^2 = -2 \sum_{i=1}^n \ln(p_i) \quad (10.6)$$

has approximately a chi-square distribution with $2n$ degrees of freedom.

This result can be used to test whether the collective results of multiple, independent hypothesis tests indicate an overall outcome that would be unlikely to occur if each of the individual hypotheses were true.

Example 10.6 Consider the situation described in Example 10.5. The significance levels from Table A3 associated with the five standard normal variables given in that example are 0.072, 0.066, 0.055, 0.082, and 0.062. By applying Equation 10.6,

$$\chi^2 = -2[\ln(0.072) + \ln(0.066) + \ln(0.055) + \ln(0.082) + \ln(0.062)] = 27.1$$

This chi-square value has 10 degrees of freedom. The value $\chi_{0.95}^2(10) = 18.3$ is obtained from Table A4. Because $27.1 > 18.3$, we conclude that the individual outcomes of all five tests should not have been so extreme (near statistical significance) if the individual hypotheses of randomness were true. This is the same conclusion reached in Example 10.5.

10.3 MULTIVARIATE STATISTICAL METHODS

Multivariate analysis is concerned with statistical procedures that are used when multiple variables are measured on each of a number of experimental units. These general procedures are analogous to univariate statistical tests and allow for a correlation structure among the multiple variables. A comparison of some multivariate methods and other statistical techniques for use in certain nuclear material accounting situations is given by Picard, Goldman, and Shipley (1982) and Telford (1982). In nuclear material management, an experiment might consist of taking an annual inventory, and the variables are inventory differences for each stratum within the inventory or each control area. (Typical strata of an inventory of a fuel fabrication plant might be uranium hexafluoride, uranium oxide powder, pellets, scrap powder, scrap pellets, sludges, air filters, solid waste, liquid waste, fuel rods, and fuel assemblies.) The multivariate statistical tests are used to test for significant differences between such inventory differences or to compare an inventory to some standard.

10.3.1 Hotelling's T^2 -Test

Hotelling's T^2 -test is an extension of Student's t-test (Section 4.2.2). To extend Student's t-test to a multidimensional random variable, each x_i

is replaced by a vector of measurements \underline{x}_i with p components. This vector is assumed to follow a multivariate normal distribution, discussed in Section 2.6.4.2, with a nonsingular covariance matrix. The null hypothesis is

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

and the two-sided alternative is

$$H_A : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

For example, μ_i might be the expected yearly inventory difference for the i^{th} stratum of an inventory, and the c_i 's might be zeros. Then the null hypothesis is that the expected inventory difference for each stratum is zero, and the alternative is that not all of these expected values are zero. The n replicate observations might be quarterly inventories.

The mean vector $\underline{\mu}$ and the covariance matrix $\underline{\Sigma}$ of a multivariate distribution are discussed in Section 2.6.3. Sample estimates $\underline{\bar{x}}$ and \underline{S} can be calculated from the formulas given in Section 3.6.1. Let \underline{c} represent the vector containing elements c_i . The null hypothesis is rejected if

$$T^2 = n(\underline{\bar{x}} - \underline{c})' \underline{S}^{-1} (\underline{\bar{x}} - \underline{c}) > \frac{p(n-1)}{n-p} F_{1-\alpha}(p, n-p) \quad (10.7)$$

where $F_{1-\alpha}(p, n-p)$ is obtained from Table A6. Note that $n > p$ observations are required for this test method.

Example 10.7 Consider a low enrichment fuel fabrication plant with a three strata inventory: uranium hexafluoride feed material, pellets, and scrap. Inventories are required annually, but informal inventories are taken

quarterly. Use Hotelling's T^2 -test to test the null hypothesis that the annual expected inventory difference for each stratum is zero. Use the $\alpha = 0.05$ level of significance. In this situation, there are $n = 4$ replicates and $p = 3$ strata or components. The data are given in Table 10.1.

TABLE 10.1
Quarterly Inventory Differences (kg)

| Strata | Quarter | | | |
|---------|---------|------|-------|------|
| | 1 | 2 | 3 | 4 |
| Feed | 17.2 | 15.0 | -11.7 | 7.2 |
| Pellets | -1.0 | 0.3 | 12.6 | -2.7 |
| Scrap | 4.9 | -3.1 | -3.5 | 3.6 |

The sample vector of average inventory differences for this year by stratum is

$$\bar{\underline{x}} = \begin{bmatrix} 6.9 \\ 2.3 \\ 0.5 \end{bmatrix}$$

and the sample covariance matrix is

$$\underline{S} = \begin{bmatrix} 172.58 & -81.09 & 30.50 \\ -81.09 & 48.66 & -21.34 \\ 30.50 & -21.34 & 19.31 \end{bmatrix}$$

The inverse of the sample covariance matrix is

$$\underline{S}^{-1} = \begin{bmatrix} 0.0287 & 0.0542 & 0.0146 \\ 0.0542 & 0.1423 & 0.0716 \\ 0.0146 & 0.0716 & 0.1079 \end{bmatrix}$$

The statement of the problem for this example implies that the vector \underline{c} has all zero elements. Thus, the left-hand side of Equation 10.7 is

$$T^2 = n \bar{\underline{x}}' \underline{S}^{-1} \bar{\underline{x}}$$

or

$$T^2 = 4[6.9, 2.3, 0.5] \begin{bmatrix} 0.0287 & 0.0542 & 0.0146 \\ 0.0542 & 0.1423 & 0.0716 \\ 0.0146 & 0.0716 & 0.1079 \end{bmatrix} \begin{bmatrix} 6.9 \\ 2.3 \\ 0.5 \end{bmatrix} = 16.52$$

With numerator degrees of freedom $p = 3$, denominator degrees of freedom $(n - p) = 1$, and $\alpha = 0.05$, the value 216 is obtained from Table A6. Thus, the right-hand side of Equation 10.7 is

$$\frac{p(n-1)}{(n-p)} F_{1-\alpha}(p, n-p) = \frac{3(3)}{1} (216) = 1944$$

Because $16.52 < 1944$, the null hypothesis of zero inventory differences by strata is not rejected at the 0.05 level of significance.

On occasion, the covariance matrix is available from historical data or from theoretical considerations. In such a situation, this matrix is assumed known rather than estimated. Then the right-hand side of Equation 10.7 is replaced by the quantile of a chi-square distribution for the desired significance level with p degrees of freedom. In Example 10.7, the computed T^2 value is 16.52. With three degrees of freedom and $\alpha = 0.05$, the value $\chi_{0.95}^2(3) = 7.81$ is obtained from Table A4. Note that if the covariance matrix had been assumed known rather than estimated, the null hypothesis would have been rejected, leading to the opposite conclusion from that reached in the example. This illustrates the decrease in sensitivity that results from the use of estimates of the covariance matrix based on small samples.

A significant T^2 statistic gives no indication of which of the components (strata) led to the rejection of the null hypothesis. Inherent in Hotelling's T^2 -test is a way of controlling the simultaneous significance level for all possible tests on linear combinations of means. Linear combinations are described by use of a vector of coefficients $\underline{a}' = [a_1, a_2, \dots, a_p]$. In Example 10.7, the second stratum can be isolated for hypothesis testing by using the coefficients $[0, 1, 0]$, and the sum of all strata inventory differences can be tested using the coefficients $[1, 1, 1]$. The simultaneous tests of all possible hypotheses of the form

$$H_0: \underline{a}'\mu = \underline{a}'c$$

$$H_A: \underline{a}'\mu \neq \underline{a}'c$$

can be performed by comparing the absolute value of the statistic

$$t(\underline{a}) = \frac{\underline{a}'(\bar{\underline{x}} - \underline{c})\sqrt{n}}{\sqrt{\underline{a}'\underline{S}\underline{a}}} \quad (10.8)$$

with the critical value

$$T_{1-\alpha}(p, n-p) = \sqrt{\frac{p(n-1)}{n-p} F_{1-\alpha}(p, n-p)}$$

If $t(\underline{a})$ exceeds the critical value, the null hypothesis is rejected at the α level of significance.

Example 10.8 Using the data from Example 10.7, test the hypothesis that the sum of all strata inventory differences is zero. In this case, \underline{a} is a vector of ones and \underline{c} is a vector of zeros. The value $T_{0.95}^2(3,1) = [(3)(3)/(1)] F_{0.95}(3,1) = 1944$ is computed in Example 10.7; thus, $T_{1-\alpha} = \sqrt{1944} = 44.1$. The denominator of Equation 10.8 is computed as

$$\underline{a}'\underline{S}\underline{a} = [1, 1, 1] \begin{bmatrix} 172.58 & -81.09 & 30.50 \\ -81.09 & 48.66 & -21.34 \\ 30.50 & -21.34 & 19.31 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 96.69$$

Then

$$t(\underline{a}) = \frac{[1, 1, 1] \begin{bmatrix} 6.9 \\ 2.3 \\ 0.5 \end{bmatrix} \sqrt{4}}{\sqrt{96.69}} = \frac{19.4}{9.83} = 1.97$$

Because $1.97 < 44.1$, the null hypothesis is not rejected at the 0.05 level of significance.

Because Equation 10.8 is a simultaneous test for all possible choices of \underline{a} , the components of \underline{a} can be chosen after the data are examined. If one

of the strata in Example 10.7 had yielded a negative inventory difference, the total absolute inventory difference would be computed by assigning a value of -1 rather than 1 to the corresponding element of \underline{a} . To test the hypothesis that the total absolute inventory difference does not exceed 5 kg, set $\underline{a}'\underline{c} = 5$. Because $\underline{a}'(\bar{\underline{x}} - \underline{c}) = \underline{a}'\bar{\underline{x}} - \underline{a}'\underline{c}$, it is not necessary to assign any values to the components of \underline{c} , and the numerator of Equation 10.8 becomes $(\underline{a}'\bar{\underline{x}} - 5)\sqrt{n}$.

10.3.2 Extensions of Hotelling's T^2 -Test

Hotelling's statistic given by Equation 10.8 can be extended to give multivariate confidence regions for means. Another extension allows two multivariate mean vectors to be compared. The null hypothesis for this extension is $H_0: \mu_1 = \mu_2$. This version of Hotelling's statistic might be used to compare inventory differences (such as those described in Example 10.7) for changes between years. The extension to more than two samples results in the one-way multivariate analysis of variance, which might be used to compare inventory differences for many years. Samples from several years at each of several facilities could be analyzed using a two-way multivariate analysis of variance. For the details of these methods, the reader is referred to texts on multivariate statistical methods, such as Morrison (1976), Kshirsagar (1972), and Anderson (1984).

10.4 TESTING FOR OUTLIERS

Sometimes there are observations present in a sample that appear to be inconsistent with other sample observations. Such values may be an extreme manifestation of the randomness inherent in the data; they may be the result of such mistakes as deviations from protocol, calibration errors, computational errors, or errors in transcribing data; or they may be an indication of uncommon factors of practical importance, such as malfunctioning equipment, material diversion, or altered records. A data analyst is sometimes tempted to discard apparently erroneous values because the loss in precision caused by removing a few good values is small compared to the loss in accuracy that may be caused by keeping even one bad value. This section provides methods that introduce statistical objectivity into identifying inconsistent data values.

The subject of statistical methods for identifying outliers has received much attention in the published literature. Some of the more recent publications include a comprehensive survey paper by Beckman and Cook (1983), books by Barnett and Lewis (1985) and Hawkins (1980), and sections of books by Kinnison (1985) and Huber (1981). The methods presented in this section are only a small subset of those found in the published literature. For additional methods and more detailed presentations, the reader should consult the above references.

Statistical tests for outliers generally have three uses: (1) to screen data routinely in preparation for analysis, (2) to identify the need for a detailed study of the process that generated the data, and (3) to find observations that may be of special interest just because they are extreme. The analyst is faced with two problems. The first is how to ascertain whether apparently discrepant values are aberrations, or **outliers**, rather than an inherent manifestation of random variation in the data. The second is to ascertain how much effect deletion or inclusion of suspected observations will have on statistical results and inferences.

Objective statistical procedures are needed for several reasons, including the following:

1. Retaining a value that is aberrant may exert a dominating, and potentially nullifying, effect on the statistical analysis.

2. Values incorrectly determined to be aberrant may contain valuable information, and their deletion may weaken or bias results.

3. Existence of observations indicated to be outliers by statistical tests may be due to incorrect assumptions about data structure rather than aberrations in the data.

4. An outlier may represent the most meaningful aspect of the data, and an objective basis for isolating such information is important. Examples are an inventory difference from a set of inventory differences that is significantly large, or a response to a cancer treatment from a battery of treatments that shows a result significantly better than the rest. In this sense every statistically significant result is an outlier.

5. The Type I error rate, the probability of incorrectly concluding that a data value is an outlier, should be controlled to a small and pre-assigned value.

In most outlier tests, the errors are assumed to be normally distributed. The normality assumption is based to some extent on habit and tradition, but there is also theoretical support based on the Central Limit Theorem and empirical support from experience with measurement results. If the normality assumption is invalid, the number of false positives (Type I errors) may be significantly increased. Skewed distributions characteristically yield excess data values that are either large or small. The lognormal is a skewed distribution that yields more small than large values. For the lognormal case, the outlier test is performed on the logarithms of the data values. In other nonnormal cases, normality may be exactly or approximately introduced by transforming the observations. Robust and non-parametric methods are available in the literature for detecting outliers when no transformation to approximate normality can be found or when the data are from a discrete distribution. There are also important outlier tests based on the uniform distribution. For the details of these and other outlier detection techniques, the reader should refer to Barnett and Lewis (1985), Beckman and Cook (1983), and Hawkins (1980).

Statistical methods for detecting outliers involve simultaneous inference because such methods are inherently *a posteriori*; that is, the need for such a method is determined after examining the data set and the same data set might be used repeatedly to test for outliers. As noted in previous sections of this chapter, the Type I error rate must be adjusted to take this into account. A wide variety of outlier detection methods is available in the statistical literature. Many of these are devised for special problems, such as outliers from a regression line. The material covered in the following sections includes: (1) the simplified outlier tests of Dixon (1950), (2) a set of tests taken from the literature and presented in an article by Grubbs (1969), and (3) tests based on studentized residuals by Beckman and Trussel (1974) which are a generalization of the earlier work of Thompson (1935). The first two sections are based on the model $y_i = \mu + \epsilon_i$, which considers an observation to consist of the sum of an expected value and a random error. The last section is based on the general linear regression model introduced in Section 6.3.1.

10.4.1 Methods of Dixon

One of the most commonly used procedures for identifying outliers is attributed to Dixon (1950). The tests are based on order statistics, and their properties are derived by using the theory of order statistics from a normal distribution. If a sample of size n is arranged in ascending order so that $x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(n)}$, then $x^{(i)}$ is called the i^{th} order statistic in the sample. In particular, $x^{(1)}$ and $x^{(n)}$ are the smallest and largest observations in the sample. The most sensitive test statistics for determining whether either the smallest or largest value is an outlier are as follows for different sample sizes:

| n | Smallest | Largest | Designation |
|-------|---|---|-------------|
| 3-7 | $(x^{(2)} - x^{(1)}) / (x^{(n)} - x^{(1)})$ | $(x^{(n-1)} - x^{(n)}) / (x^{(1)} - x^{(n)})$ | r_{10} |
| 8-10 | $(x^{(2)} - x^{(1)}) / (x^{(n-1)} - x^{(1)})$ | $(x^{(n-1)} - x^{(n)}) / (x^{(2)} - x^{(n)})$ | r_{11} |
| 11-13 | $(x^{(3)} - x^{(1)}) / (x^{(n-1)} - x^{(1)})$ | $(x^{(n-2)} - x^{(n)}) / (x^{(2)} - x^{(n)})$ | r_{21} |
| 14-25 | $(x^{(3)} - x^{(1)}) / (x^{(n-2)} - x^{(1)})$ | $(x^{(n-2)} - x^{(n)}) / (x^{(3)} - x^{(n)})$ | r_{22} |

Critical values for these test statistics by sample size are given in Table 10.2. For the column headed by $P_{100(1-\alpha)}$, the probability is α that the test statistic will exceed the tabulated value, under the assumption that the sample comes from a normal distribution.

To excise all aberrant values, Dixon (1950) suggests repeated use of the test statistics r_{10} , r_{11} , r_{21} , or r_{22} , treating the first suspected observation

TABLE 10.2
Criteria for Testing for Extreme Values†

| Statistic | n | P ₇₀ | P ₈₀ | P ₉₀ | P ₉₅ | P ₉₈ | P ₉₉ | P _{99.5} |
|-----------------|----|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------------------|
| r ₁₀ | 3 | 0.684 | 0.781 | 0.886 | 0.941 | 0.976 | 0.988 | 0.994 |
| | 4 | 0.471 | 0.560 | 0.679 | 0.765 | 0.846 | 0.889 | 0.926 |
| | 5 | 0.373 | 0.451 | 0.557 | 0.642 | 0.729 | 0.780 | 0.821 |
| | 6 | 0.318 | 0.386 | 0.482 | 0.560 | 0.644 | 0.698 | 0.740 |
| | 7 | 0.281 | 0.344 | 0.434 | 0.507 | 0.586 | 0.637 | 0.680 |
| r ₁₁ | 8 | 0.318 | 0.385 | 0.479 | 0.554 | 0.631 | 0.683 | 0.725 |
| | 9 | 0.288 | 0.352 | 0.441 | 0.512 | 0.587 | 0.635 | 0.677 |
| | 10 | 0.265 | 0.325 | 0.409 | 0.477 | 0.551 | 0.597 | 0.639 |
| r ₂₁ | 11 | 0.391 | 0.442 | 0.517 | 0.576 | 0.638 | 0.679 | 0.713 |
| | 12 | 0.370 | 0.419 | 0.490 | 0.546 | 0.605 | 0.642 | 0.675 |
| | 13 | 0.351 | 0.399 | 0.467 | 0.521 | 0.578 | 0.615 | 0.649 |
| r ₂₂ | 14 | 0.370 | 0.421 | 0.492 | 0.546 | 0.602 | 0.641 | 0.674 |
| | 15 | 0.353 | 0.402 | 0.472 | 0.525 | 0.579 | 0.616 | 0.647 |
| | 16 | 0.338 | 0.386 | 0.454 | 0.507 | 0.559 | 0.595 | 0.624 |
| | 17 | 0.325 | 0.373 | 0.438 | 0.490 | 0.542 | 0.577 | 0.605 |
| | 18 | 0.314 | 0.361 | 0.424 | 0.475 | 0.527 | 0.561 | 0.589 |
| | 19 | 0.304 | 0.350 | 0.412 | 0.462 | 0.514 | 0.547 | 0.575 |
| | 20 | 0.295 | 0.340 | 0.401 | 0.450 | 0.502 | 0.535 | 0.562 |
| | 21 | 0.287 | 0.331 | 0.391 | 0.440 | 0.491 | 0.524 | 0.551 |
| | 22 | 0.280 | 0.323 | 0.382 | 0.430 | 0.481 | 0.514 | 0.541 |
| | 23 | 0.274 | 0.316 | 0.374 | 0.421 | 0.472 | 0.505 | 0.532 |
| | 24 | 0.268 | 0.310 | 0.367 | 0.413 | 0.464 | 0.497 | 0.524 |
| | 25 | 0.262 | 0.304 | 0.360 | 0.406 | 0.457 | 0.489 | 0.516 |

†Reproduced from Dixon (1953), with permission from the Biometric Society.

as being from a sample of n observations, the second as being from a sample of $n - 1$ observations, and so on. In practice, this procedure does not necessarily eliminate the same extremes or even the same number of extremes as other techniques. This is not surprising, because the process only guarantees that under the null hypothesis, values that are not really outliers (false positives) will be excised $100\alpha\%$ of the time on the average. The various outlier tests tend to handle alternative hypotheses differently, so the same observations may not be judged to be outliers when different tests are applied. In particular, all values tested to be aberrant may be eliminated in one step if the set of the smallest number of values to form a homogeneous group is correctly gauged.

It should be noted that if Dixon's test method is applied more than once to a single set of data, then there is simultaneous inference, and the

overall significance level is larger than that specified for each individual test. Tietjen and Moore (1972) address the problem of repeated application of a test method to a single set of data. They also present techniques for simultaneously testing either the k largest values or the k most extreme (largest and/or smallest) values with a specified level of significance.

Example 10.9 Ten ordered observations of cladding thickness (in inches) on fuel rods are 0.0284, 0.0285, 0.0285, 0.0285, 0.0286, 0.0286, 0.0286, 0.0289, 0.0292, and 0.0298. The suspicious observation is $x^{(10)} = 0.0298$, the highest value. Because $n = 10$, compute

$$r_{11} = \frac{0.0292 - 0.0298}{0.0284 - 0.0298} = 0.429$$

From Table 10.2, the probability that a value of r_{11} greater than 0.429 would occur by chance is about 0.08, so this observation would be rejected at the 0.10 level of significance but not at the 0.05 level. Following recommendations given by Dixon (1953), this value would be eliminated if the presence of aberrant observations was a relatively common occurrence. In this case, it might be desirable to continue by testing the largest of the remaining 9 observations by computing

$$r_{11} = \frac{0.0289 - 0.0292}{0.0284 - 0.0292} = 0.375$$

This value is not significant at the 0.10 level of significance, so no further observations would be eliminated, even if aberrant observations are known to occur frequently.

10.4.2 Grubbs' Treatment of Outliers

Grubbs (1969) published a comprehensive article on outliers. The cases covered involve testing the largest, the smallest, the two largest, the two smallest, and the largest and smallest observations together, as well as repeated testing at one end or both ends of an ordered sample. Cases are also covered where the degrees of freedom in the estimated standard deviation exceed the sample size and for the special limiting case where the standard deviation is known. The methods of Dixon and the associated tables are included and discussed by Grubbs (1969). The statistics in Table 10.3 occur in one or more of the cases.

TABLE 10.3
Statistics Used to Test for Outliers

| Statistic | Description |
|---|--|
| $\bar{x} = \sum_{i=1}^n x_i/n$ | Sample average (Equation 1.1). |
| $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \left[\sum_{i=1}^n x_i \right]^2/n}{n-1}$ | Sample variance (Equation 1.8). |
| $s = \sqrt{s^2}$ | Sample standard deviation. |
| $SS = (n-1)s^2$ | The sample sum of squared deviations (SSSD). |
| $SS_{1,2} = \sum_{i=3}^n x_i^2 - \left[\sum_{i=3}^n x_i \right]^2/(n-2)$ | The SSSD omitting the two smallest observations. |
| $SS_{n-1,n} = \sum_{i=1}^{n-2} x_i^2 - \left[\sum_{i=1}^{n-2} x_i \right]^2/(n-2)$ | The SSSD omitting the two largest observations. |
| $\sqrt{b_1} = \frac{\sqrt{n} \sum_{i=1}^n (x_i - \bar{x})^3}{[s^2(n-1)]^{3/2}}$ | Sample coefficient of skewness. |
| $b_2 = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)^2 s^4}$ | Sample coefficient of kurtosis. |
| $w = x^{(n)} - x^{(1)}$ | Sample range of values. |

Outlier problems that are handled by Grubbs (1969) and the corresponding recommended approaches are given in Table 10.4. Grubbs (1969) makes some important and insightful remarks on outliers. In particular,

When the skilled experimenter is clearly aware that a gross deviation from prescribed experimental procedure has taken place, the resultant observations (sic) should be discarded, whether or not it agrees with the rest of the data and without recourse to statistical tests for outliers. If a reliable correction procedure, for example, for temperature, is available, the observation may sometimes be corrected and retained.

Although our primary interest here is that of detecting outlying observations, we remark that the statistical criteria used also test the hypothesis that the random sample taken did indeed come from a normal or Gaussian population.

TABLE 10.4
Outlier Problems and Appropriate Test Statistics

| Problem | Statistic | Comment |
|---|-------------------------------------|--|
| Testing the largest value as an outlier | $T_n = \frac{x^{(n)} - \bar{x}}{s}$ | Critical values are provided at the 0.01, 0.025, and 0.05 significance levels, for $n = 3(1)25$, $n = 25(5)50$, and $n = 50(10)100$. These are given in Table A21. If there is no <i>a priori</i> reason for testing either a largest or smallest value as an outlier, the actual significance levels are 0.02, 0.05, and 0.10. |
| Testing the smallest value as an outlier | $T_1 = \frac{\bar{x} - x^{(1)}}{s}$ | |
| Testing jointly the two extremes as outliers | w/s | Critical values for the ratio of the sample range and standard deviation are given in Table A22 for the 0.005, 0.01, and 0.05 significance levels. A value of w/s that exceeds a critical value is statistically significant at that level. If significance is found, the smallest and largest values should be tested to determine the reason for significance. |
| Testing jointly the two smallest observations as outliers | $\frac{SS_{1,2}}{SS}$ | Small values indicate significance. Critical values are found in Table A23 for the 0.01, 0.05, and 0.10 significance levels. |
| Testing jointly the two largest observations as outliers | $\frac{SS_{n-1,n}}{SS}$ | |

| | | |
|--|--|---|
| Repeated testing at either the largest or the smallest end for outliers. | $\sqrt{b_1}$ | This statistic should be recalculated for the reduced n value after each significant outlier is eliminated from the data. Critical values are found in Table A24 for the 0.01 and 0.05 levels for $n = 5(5)50$ and $n = 60$. |
| Repeated testing at both ends for outliers. | b_2 | This statistic should be recalculated for the reduced n after each significant outlier is eliminated from the data. Critical values are found in Table A25 for $n = 5(5)25$, $n = 25(25)100$ for the 0.01 and 0.05 levels. |
| Testing extremes when there are extra degrees of freedom in the standard deviation estimate. | $T'_n = \frac{x^{(n)} - \bar{x}}{s_\nu}$ $T'_1 = \frac{\bar{x} - x^{(1)}}{s_\nu}$ $s_\nu^2 = \frac{\sum \nu_k s_k^2}{\sum \nu_k}$ $\nu = \sum \nu_k$ | The case considered here is when the precision of the variance estimate is enhanced by pooling with previous variance estimates. The pooled estimate is a weighted average of the previous estimates (weighted by their degrees of freedom). Critical values are found in Table A26 for the 0.01 and 0.05 significance levels for $n = 3(1)10$, and 12, with $\hat{\nu}$ = degrees of freedom = 10(1)20, $\nu = 24, 30, 40, 60, 120$. For the case where σ is known (i.e., $\nu = \infty$), critical values are found in Table A27 for the 0.005, 0.01, and 0.05 significance levels for $n = 2(1)25$. |

The calculation of the critical points for statistics such as T_1 , T_n , $\sqrt{b_1}$, b_2 , w/s , $SS_{1,2}$, and $SS_{n-1,n}$ involves the theory of order statistics from a normal distribution. Conceptually, the processes for determining critical values are not difficult. Computations can be difficult, however, and the solutions can require numerical methods or Monte Carlo methods (Ferguson, 1961).

It is important to note that if more than one of the tests in Table 10.4 is applied to a set of data, or if one of the tests is applied more than once to a set of data, then there is simultaneous inference and degradation of the overall level of significance. When more than one of the tests is to be applied to a single set of data, the Bonferroni inequality (Equation 10.1) can be applied to control the overall (simultaneous) level of significance. For example, if four of the test methods are to be used, then setting the significance level of each test at $\alpha/4$ will result in an overall significance level that is less than or equal to α . As noted in Section 10.4.1, Tietjen and Moore (1972) address the problem of repeated application of a single test method to a set of data. They also introduce techniques for simultaneously testing either the k largest values or the k most extreme (largest and/or smallest) values with a specified level of significance.

Example 10.10 The statistics T_1 and T_n from Table 10.4 are frequently recommended for testing extreme observations, particularly if n is relatively large. For the data in Example 10.9, the sample mean is $\bar{x} = 0.02876$ and the sample standard deviation is $s = 0.000435$, so that

$$T_{10} = (0.0298 - 0.02876)/0.000435 = 2.39$$

Entering Table A21 and noting there was no prior reason to test the largest as opposed to the extreme deviation in either direction, this value is significant at almost the $2(0.01) = 0.02$ level for $n = 10$ and is apparently more aberrant than was indicated by the Dixon test based on ratios. However, the Dixon test is less affected by lack of symmetry in the data.

10.4.3 Studentized Residual Approach

A general procedure for assessing whether one or more of the observations in a least squares regression are outliers is given by Beckman and Trussel (1974). The method is based on studentized residuals and has the following useful characteristics:

1. It is of wide generality.
2. It includes some other methods, such as those of Thompson (1935) and Grubbs (1969), as special cases.
3. It can be computed rather easily using results from standard regression software.

The procedure is particularly useful in pinpointing outliers in inter-laboratory experiments when the analysis is based on a linear model. Obtaining good estimates of the variances of random errors of measurement is very important for safeguards data. Experience has shown that outliers seriously and adversely affect the process of estimating measurement variances. The method presented in this section is effective in eliminating outliers and thus enhancing the variance estimation process.

The general linear regression model is defined in Section 6.3.1. The computational methods for estimating the coefficients in the model are given in Section 6.3.2. The standardized residuals r_i^* and the elements h_{ij} of the hat matrix are defined in Section 6.4.5.2 for a model that has been fitted to a set of data. Methods for plotting and approximate methods for statistically testing the standardized residuals to identify outliers in regression data are presented in Section 6.4.5.2.

Another approach to examining residuals is to eliminate the i^{th} observation from the fitting process and then determine the i^{th} standardized residual, so the estimated residual is independent of the observation. Because the observation and the variance estimate are independent, the resulting **studentized residuals** η_i have a Student's t -distribution with $n - p - 1$ degrees of freedom under the null hypothesis. This process would impose a computational burden if it were not for the fact that the η_i and r_i^* are related by the identity (see Beckman and Trussel, 1974).

$$\eta_i = r_i^* \left(\frac{n - p - 1}{n - p - r_i^{*2}} \right)^{1/2} \quad (10.9)$$

When testing n studentized residuals, the overall Type I error rate should be controlled by using methods of simultaneous inference. Critical values corresponding to an error rate $\alpha' = 1 - (1 - \alpha)^{1/n}$, which for an individual test will ensure an overall error rate α , usually cannot be determined directly from tables of Student's t -distribution. To facilitate screening for outliers on residuals obtained when fitting a linear model containing p parameters to n observations, Table A28 gives adjusted critical points for η_i computed from the approximation to the Student's t -distribution presented by Fisher and Cornish (1960). Because this approximation allows the computation of $t_{1 - \alpha} (df)$ directly from $z_{1 - \alpha}$ of the

standard normal distribution, the approximation is convenient for use in computerized programs that include screening for outliers in routine processing.

In summary, the method of using studentized residuals to test for outliers in regression involves the following steps:

1. Obtain the usual regression results.
2. Calculate the h_{ij} values, or the H matrix.
3. Determine the predicted response variables \hat{y}_i , the residuals $r_i = y_i - \hat{y}_i$, and the standardized residuals r_i^* .
4. Compute the studentized residuals η_i , which have a Student's *t*-distribution under the null hypothesis of (a) a correct structure for the expected value in the regression model, (b) independent errors that are normally distributed, and (c) no outliers.
5. Compare $|\eta_i|$ with the appropriate critical value from Table A28 to test for outliers.

As with previous procedures, the process should be iterated until, at the final iteration, no outliers are found.

Example 10.11 The following data represent determinations of the plutonium content of each of 10 samples by four laboratories:

| Sample | Laboratory | | | | Sample average |
|--------------------|------------|---------|---------|---------|----------------|
| | 1 | 2 | 3 | 4 | |
| 1 | 1.32300 | 1.38000 | 1.36900 | Missing | 1.35733 |
| 2 | 1.37400 | 1.40300 | 1.39600 | Missing | 1.39100 |
| 3 | 1.35400 | 1.39000 | 1.23000 | 1.37000 | 1.33600 |
| 4 | 1.37300 | 1.39500 | 1.38400 | Missing | 1.38400 |
| 5 | 1.70000 | 1.60100 | 1.58400 | 1.60700 | 1.62300 |
| 6 | 1.51000 | 1.54000 | 1.51600 | 1.56700 | 1.53325 |
| 7 | 1.42300 | 1.46100 | 1.45700 | 1.46500 | 1.45150 |
| 8 | 1.39800 | 1.43800 | 1.44200 | 1.46300 | 1.43525 |
| 9 | 1.41000 | 1.45700 | 1.42600 | 1.43700 | 1.43250 |
| 10 | 1.29500 | 1.32700 | 1.32400 | 1.41000 | 1.33900 |
| Laboratory average | 1.41600 | 1.43920 | 1.41280 | 1.47414 | |

This is a randomized block design (Section 7.2.2), where the samples are blocks and the laboratories are treatments. Because of three missing values, however, least squares regression methods were used to fit a linear

model with 13 parameters to the 37 observations. The 13 parameters include a constant value, 3 laboratory effects, and 9 sample effects. Least squares regression techniques for fitting analysis of variance models involve some special considerations (primarily concerned with defining the \mathbf{X} matrix and the β vector) that are not presented in this book. The reader is referred to Chapter 9 of Draper and Smith (1981) for details.

The predicted values, residuals, standardized residuals, and studentized residuals from the fitted model are displayed in Table 10.5. From Table

TABLE 10.5
Modeling Results with All 37 Data Points Included

| No. | Sample | Lab | y_i | \hat{y}_i | $y_i - \hat{y}_i$ | r_i^* | η_i |
|-----|--------|-----|-------|-------------|-------------------|---------|----------|
| 1 | 1 | 1 | 1.323 | 1.35067 | -0.02767 | -1.00 | -1.00 |
| 2 | 1 | 2 | 1.380 | 1.37387 | 0.00613 | 0.22 | 0.22 |
| 3 | 1 | 3 | 1.369 | 1.34747 | 0.02153 | 0.78 | 0.77 |
| 4 | 2 | 1 | 1.374 | 1.38433 | -0.01333 | -0.37 | -0.37 |
| 5 | 2 | 2 | 1.403 | 1.40753 | -0.00453 | -0.16 | -0.16 |
| 6 | 2 | 3 | 1.396 | 1.38113 | 0.01487 | 0.54 | 0.53 |
| 7 | 3 | 1 | 1.354 | 1.32131 | 0.03269 | 1.12 | 1.12 |
| 8 | 3 | 2 | 1.390 | 1.34451 | 0.04549 | 1.55 | 1.60 |
| 9 | 3 | 3 | 1.230 | 1.31811 | -0.08811 | -3.01 | -3.73 |
| 10 | 3 | 4 | 1.370 | 1.36007 | 0.00993 | 0.35 | 0.34 |
| 11 | 4 | 1 | 1.373 | 1.37733 | -0.00433 | -0.16 | -0.15 |
| 12 | 4 | 2 | 1.395 | 1.40053 | -0.00553 | -0.20 | -0.20 |
| 13 | 4 | 3 | 1.384 | 1.37413 | 0.00987 | 0.36 | 0.35 |
| 14 | 5 | 1 | 1.700 | 1.60831 | 0.09169 | 3.13 | 3.99 |
| 15 | 5 | 2 | 1.601 | 1.63151 | -0.03051 | -1.04 | -1.04 |
| 16 | 5 | 3 | 1.584 | 1.60511 | -0.02111 | -0.72 | -0.71 |
| 17 | 5 | 4 | 1.607 | 1.64707 | -0.04007 | -1.40 | -1.43 |
| 18 | 6 | 1 | 1.510 | 1.51856 | -0.00856 | -0.29 | -0.29 |
| 19 | 6 | 2 | 1.540 | 1.54176 | -0.00176 | -0.06 | -0.06 |
| 20 | 6 | 3 | 1.516 | 1.51536 | 0.00064 | 0.02 | 0.02 |
| 21 | 6 | 4 | 1.567 | 1.55732 | 0.00968 | 0.34 | 0.33 |
| 22 | 7 | 1 | 1.423 | 1.43681 | -0.01381 | -0.47 | -0.46 |
| 23 | 7 | 2 | 1.461 | 1.46001 | 0.00099 | 0.03 | 0.03 |
| 24 | 7 | 3 | 1.457 | 1.43361 | 0.02339 | 0.80 | 0.79 |
| 25 | 7 | 4 | 1.465 | 1.47557 | -0.01057 | -0.37 | -0.36 |
| 26 | 8 | 1 | 1.398 | 1.42056 | -0.02256 | -0.77 | -0.76 |
| 27 | 8 | 2 | 1.438 | 1.44376 | -0.00576 | -0.20 | -0.19 |
| 28 | 8 | 3 | 1.442 | 1.41736 | 0.02464 | 0.84 | 0.84 |
| 29 | 8 | 4 | 1.463 | 1.45932 | 0.00368 | 0.13 | 0.13 |
| 30 | 9 | 1 | 1.410 | 1.41781 | -0.00781 | -0.27 | -0.26 |
| 31 | 9 | 2 | 1.457 | 1.44101 | 0.01499 | 0.55 | 0.54 |
| 32 | 9 | 3 | 1.426 | 1.41461 | 0.01139 | 0.39 | 0.38 |
| 33 | 9 | 4 | 1.437 | 1.45657 | -0.01957 | -0.68 | -0.68 |
| 34 | 10 | 1 | 1.295 | 1.32431 | -0.02931 | -1.00 | -1.00 |
| 35 | 10 | 2 | 1.327 | 1.34751 | -0.02051 | -0.70 | -0.69 |
| 36 | 10 | 3 | 1.324 | 1.32111 | 0.00289 | 0.10 | 0.10 |
| 37 | 10 | 4 | 1.410 | 1.36307 | 0.04693 | 1.64 | 1.70 |

A28 with $n = 37$ and $p = 13$, the critical value for η_i at the 0.05 significance level is 3.64. Thus, the underlined observations 9 and 14 have significant η_i values and are suspected of being aberrant. Note that the third largest residual is less than $\frac{1}{2}$ of the critical value.

After eliminating points 9 and 14 from the data, the model was refitted to the 35 remaining points and Table 10.6 was produced. The value $\eta_{35} = 5.52$ exceeds 3.66, the critical value at the 0.05 significance level. By eliminating this observation and refitting the model to the 34 remaining points,

TABLE 10.6
Modeling Results with Points 9 and 14 Removed

| No. | Sample | Lab | y_i | \hat{y}_i | $y_i - \hat{y}_i$ | r_i^* | η_i |
|-----|--------|-----|-------|-------------|-------------------|---------|----------|
| 1 | 1 | 1 | 1.323 | 1.33675 | -0.01375 | -1.10 | -1.11 |
| 2 | 1 | 2 | 1.380 | 1.37404 | 0.00596 | 0.48 | 0.47 |
| 3 | 1 | 3 | 1.369 | 1.36121 | 0.00779 | 0.62 | 0.62 |
| 4 | 2 | 1 | 1.374 | 1.37042 | 0.00358 | 0.29 | 0.28 |
| 5 | 2 | 2 | 1.403 | 1.40771 | -0.00471 | -0.38 | -0.37 |
| 6 | 2 | 3 | 1.396 | 1.39488 | 0.00112 | 0.09 | 0.09 |
| 7 | 3 | 1 | 1.354 | 1.34126 | 0.01274 | 1.03 | 1.03 |
| 8 | 3 | 2 | 1.390 | 1.37855 | 0.01145 | 0.92 | 0.91 |
| 9 | 3 | 4 | 1.370 | 1.39419 | -0.02419 | -1.97 | -2.13 |
| 10 | 4 | 1 | 1.373 | 1.36342 | 0.00958 | 0.77 | 0.76 |
| 11 | 4 | 2 | 1.395 | 1.40071 | -0.00571 | -0.46 | -0.45 |
| 12 | 4 | 3 | 1.384 | 1.38788 | -0.00388 | -0.31 | -0.30 |
| 13 | 5 | 2 | 1.601 | 1.59640 | 0.00460 | 0.37 | 0.36 |
| 14 | 5 | 3 | 1.584 | 1.58357 | 0.00043 | 0.03 | 0.03 |
| 15 | 5 | 4 | 1.607 | 1.61203 | -0.00503 | -0.41 | -0.40 |
| 16 | 6 | 1 | 1.510 | 1.50458 | 0.00542 | 0.41 | 0.40 |
| 17 | 6 | 2 | 1.540 | 1.54187 | -0.00187 | -0.14 | -0.14 |
| 18 | 6 | 3 | 1.516 | 1.52904 | -0.01304 | -0.99 | -0.99 |
| 19 | 6 | 4 | 1.567 | 1.55751 | 0.00949 | 0.73 | 0.72 |
| 20 | 7 | 1 | 1.423 | 1.42283 | 0.00017 | 0.01 | 0.01 |
| 21 | 7 | 2 | 1.461 | 1.46012 | 0.00088 | 0.07 | 0.06 |
| 22 | 7 | 3 | 1.457 | 1.44729 | 0.00971 | 0.74 | 0.73 |
| 23 | 7 | 4 | 1.465 | 1.47576 | -0.01076 | -0.83 | -0.82 |
| 24 | 8 | 1 | 1.398 | 1.40658 | -0.00858 | -0.65 | -0.64 |
| 25 | 8 | 2 | 1.438 | 1.44387 | -0.00587 | -0.44 | -0.43 |
| 26 | 8 | 3 | 1.442 | 1.43104 | 0.01096 | 0.83 | 0.83 |
| 27 | 8 | 4 | 1.463 | 1.45951 | 0.00349 | 0.27 | 0.26 |
| 28 | 9 | 1 | 1.410 | 1.40383 | 0.00617 | 0.47 | 0.46 |
| 29 | 9 | 2 | 1.457 | 1.44112 | 0.01588 | 1.20 | 1.21 |
| 30 | 9 | 3 | 1.426 | 1.42829 | -0.00229 | -0.17 | -0.17 |
| 31 | 9 | 4 | 1.437 | 1.45676 | -0.01976 | -1.52 | -1.17 |
| 32 | 10 | 1 | 1.295 | 1.31033 | -0.01533 | -1.16 | -1.17 |
| 33 | 10 | 2 | 1.327 | 1.34762 | -0.02062 | -1.55 | -1.61 |
| 34 | 10 | 3 | 1.324 | 1.33479 | -0.01079 | -0.82 | -0.81 |
| 35 | 10 | 4 | 1.410 | 1.35326 | 0.04674 | 3.61 | 5.52 |

Table 10.7 was produced, in which no studentized residual exceeds 3.67, the critical value at the 0.05 significance level.

The effect of this screening on the residual sum of squares and the significance of the fit can be seen by examining the analysis of variance tables displayed in Table 10.8, which also show the additional variance explained by each successive set of effects and the residual variance. Note that because of missing values, the variance explained by laboratories and samples is dependent on the order in which the effects are estimated.

TABLE 10.7
Modeling Results with 3 Outliers Removed

| No. | Sample | Lab | y_i | \hat{y}_i | $y_i - \hat{y}_i$ | r_i^* | η_i |
|-----|--------|-----|-------|-------------|-------------------|---------|----------|
| 1 | 1 | 1 | 1.323 | 1.33688 | -0.01388 | -1.70 | -1.79 |
| 2 | 1 | 2 | 1.380 | 1.37379 | 0.00621 | 0.76 | 0.75 |
| 3 | 1 | 3 | 1.369 | 1.36134 | 0.00766 | 0.94 | 0.94 |
| 4 | 2 | 1 | 1.374 | 1.37054 | 0.00346 | 0.42 | 0.42 |
| 5 | 2 | 2 | 1.403 | 1.40745 | -0.00445 | -0.54 | -0.53 |
| 6 | 2 | 3 | 1.396 | 1.39500 | 0.00100 | 0.12 | 0.12 |
| 7 | 3 | 1 | 1.354 | 1.34503 | 0.00897 | 1.11 | 1.12 |
| 8 | 3 | 2 | 1.390 | 1.38194 | 0.00806 | 0.99 | 0.99 |
| 9 | 3 | 4 | 1.370 | 1.38702 | -0.01702 | -2.15 | -2.38 |
| 10 | 4 | 1 | 1.373 | 1.36354 | 0.00946 | 1.16 | 1.17 |
| 11 | 4 | 2 | 1.395 | 1.40045 | -0.00545 | -0.67 | -0.66 |
| 12 | 4 | 3 | 1.384 | 1.38800 | -0.00400 | -0.49 | -0.48 |
| 13 | 5 | 2 | 1.601 | 1.59979 | 0.00121 | 0.15 | 0.15 |
| 14 | 5 | 3 | 1.584 | 1.58734 | -0.00334 | -0.41 | -0.40 |
| 15 | 5 | 4 | 1.607 | 1.60487 | 0.00213 | 0.27 | 0.26 |
| 16 | 6 | 1 | 1.510 | 1.50741 | 0.00259 | 0.30 | 0.29 |
| 17 | 6 | 2 | 1.540 | 1.54432 | -0.00432 | -0.50 | -0.49 |
| 18 | 6 | 3 | 1.516 | 1.53187 | -0.01587 | -1.84 | -1.97 |
| 19 | 6 | 4 | 1.567 | 1.54950 | 0.01750 | 2.11 | 2.32 |
| 20 | 7 | 1 | 1.423 | 1.42566 | -0.00266 | -0.31 | -0.30 |
| 21 | 7 | 2 | 1.461 | 1.46257 | -0.00157 | -0.18 | -0.18 |
| 22 | 7 | 3 | 1.457 | 1.45012 | 0.00688 | 0.80 | 0.79 |
| 23 | 7 | 4 | 1.465 | 1.46765 | -0.00265 | -0.32 | -0.31 |
| 24 | 8 | 1 | 1.398 | 1.40941 | -0.01141 | -1.33 | -1.35 |
| 25 | 8 | 2 | 1.438 | 1.44632 | -0.00832 | -0.96 | -0.96 |
| 26 | 8 | 3 | 1.442 | 1.43387 | 0.00813 | 0.94 | 0.94 |
| 27 | 8 | 4 | 1.463 | 1.45140 | 0.01160 | 1.39 | 1.42 |
| 28 | 9 | 1 | 1.410 | 1.40666 | 0.00334 | 0.39 | 0.38 |
| 29 | 9 | 2 | 1.457 | 1.44357 | 0.01343 | 1.55 | 1.61 |
| 30 | 9 | 3 | 1.426 | 1.43112 | -0.00512 | -0.60 | -0.59 |
| 31 | 9 | 4 | 1.437 | 1.44865 | -0.01165 | -1.40 | -1.43 |
| 32 | 10 | 1 | 1.295 | 1.29488 | 0.00012 | 0.02 | 0.01 |
| 33 | 10 | 2 | 1.327 | 1.33179 | -0.00479 | -0.58 | -0.58 |
| 34 | 10 | 3 | 1.324 | 1.31934 | 0.00466 | 0.57 | 0.56 |

TABLE 10.8
Analysis of Variance Results Illustrating
the Effect of Removing Outliers

| | Sum of squares | df | Mean square | F ratio |
|---------------|----------------|----|-------------|---------|
| n = 37 | | | | |
| Mean | 75.9161 | 1 | 75.9161 | |
| Samples | 0.278997 | 9 | 0.0309997 | 24.27 |
| Laboratories | 0.0191909 | 3 | 0.00639697 | 5.01 |
| Residual | 0.0306547 | 24 | 0.00127728 | |
| n = 35 | | | | |
| Mean | 71.6259 | 1 | 71.6259 | |
| Samples | 0.177128 | 9 | 0.0196809 | 74.94 |
| Laboratories | 0.0332411 | 3 | 0.0110804 | 42.19 |
| Residual | 0.00577768 | 22 | 0.000262622 | |
| n = 34 | | | | |
| Mean | 69.6382 | 1 | 69.6382 | |
| Samples | 0.175746 | 9 | 0.0195273 | 173.82 |
| Laboratories | 0.0376067 | 3 | 0.0125356 | 111.58 |
| Residual | 0.00235921 | 21 | 0.000112343 | |

The elimination of 3 of the 37 observations has reduced the residual variance by a factor of 10 and has greatly enhanced the significance of the differences among samples and laboratories. The latter is important because of the increased confidence in the estimates of the between-laboratory biases obtained from the data, as well as confidence in the ability of the method to characterize process variability.

REFERENCES

- Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Barnett, V., and T. Lewis, 1985, *Outliers in Statistical Data*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Bechhofer, R. E., and C. W. Dunnett, 1982, Multiple Comparisons for Orthogonal Contrast: Examples and Tables, *Technometrics*, **24**: 213-222.
- Beckman, R. J., and R. D. Cook, 1983, Outliers, *Technometrics*, **25**: 119-149.
- , and H. J. Trussell, 1974, The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression, *J. Am. Stat. Assoc.*, **69**: 199-201.

- Dixon, W. J., 1950, Analysis of Extreme Values, *Ann. Math. Stat.*, 21: 488-506.
- , 1953, Processing Data for Outliers, *Biometrics*, 9: 74-89.
- Draper, N. R., and H. Smith, 1981, *Applied Regression Analysis*, 2nd Ed., John Wiley & Sons, Inc., New York.
- Ferguson, T. S., 1961, Rules for Rejection of Outliers, *Revue Inst. de Stat.* 3: 29-43.
- Fisher, R. A., 1970, *Statistical Methods for Research Workers*, (Section 21.1), 14th Ed., Oliver and Boyd, Edinburgh, Scotland.
- , and E. A. Cornish, 1960, The Percentile Points of Distributions Having Known Cumulants, *Technometrics*, 2: 209-225.
- Gabriel, K. R., 1978, A Simple Method of Multiple Comparisons, *J. Am. Stat. Assoc.*, 73: 724-729.
- Grubbs, F. E., 1969, Procedures for Detecting Outlying Observations in Samples, *Technometrics*, 11: 1-21.
- Hawkins, D. M., 1980, *Identification of Outliers*, Chapman and Hall (in association with Methuen, Inc.), New York.
- Huber, P. J., 1981, *Robust Statistics*, John Wiley & Sons, Inc., New York.
- Kinnison, R. R., 1985, *Applied Extreme Value Statistics*, Macmillan Publishing Co., New York.
- Kshirsagar, A. M., 1972, *Multivariate Analysis*, Marcel Dekker, Inc., New York.
- Jaech, J. L., 1974, Testing for Normality When the Data Are Grouped Due to Rounding, *Nucl. Mater. Manage.*, 3(2): 40-46.
- Miller, R. G., 1966, *Simultaneous Statistical Inference*, McGraw-Hill, New York.
- Morrison, D. F., 1976, *Multivariate Statistical Methods*, 2nd Ed., McGraw-Hill, New York.
- Picard, R., A Goldman, and J. P. Shipley, 1982, A Comparison of Tests for Spatially Distributed Materials Balances, Part I, *Proc. 1981 DOE Statistical Symposium*, Brookhaven National Laboratory, Upton, Long Island, New York.
- Telford, J. 1982, A Comparison of Tests for Spatially Distributed Materials Balances, Part II, *Proc. 1981 DOE Statistical Symposium*, Brookhaven National Laboratory, Upton, Long Island, New York.
- Tietjen, G. L., and R. H. Moore, 1972, Some Grubbs-Type Statistics for the Detection of Several Outliers, *Technometrics*, 14: 583-597.
- Thompson, W. R., 1935, On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to Standard Deviation, *Ann. Math. Stat.*, 6: 214-219.

CHAPTER 11

Simulation

11.0 INTRODUCTION

Simulation is a science (or discipline) that includes a large class of procedures for problem solving by defining and analyzing models of complex processes or systems. Briefly, simulation involves developing a representative model of a process or system and then manipulating and experimenting with the model on a computer to improve understanding of the process or system. For example, if the model is a valid representation, it is possible by experimenting with the model to observe how changes in certain parameters might affect the operation or response of the process or system.

In nuclear material research and management, there is an emphasis on developing models to describe and analyze production, measurement, and quality control systems. Many of these models could be manipulated on a computer to investigate the effects on the response or output of the system due to changes in components of the system. Thus, simulation analysis is potentially a valuable tool for nuclear material research and management but is not currently as widely utilized as it could be.

An important tool in simulation modeling and analysis is the generation of random numbers that can be transformed to represent a random sample from a specified probability distribution. This enables the study and evaluation of complicated functions of random variables which may be difficult or impossible to investigate analytically. Simulation methods that involve generating random deviates are commonly called Monte Carlo methods. For simplicity, however, this distinction is not emphasized in this chapter.

Section 11.3 provides an introduction to generating random deviates from a few specific probability distributions. Although Section 11.3 comprises a major portion of this chapter, it is limited in scope and content. It illustrates, by simple examples, how simulation techniques that utilize random deviate generation can be applied to investigate the properties of the statistical procedures presented throughout this book. Although most of the examples presented have straightforward analytical solutions, the general concepts illustrated can be generalized and extended to aid in solving problems that are difficult (or practically impossible) to solve analytically.

It must be emphasized that this chapter is not intended to be a comprehensive guide to simulation modeling and analysis. Its purpose is only

to bring simulation to the attention of the practitioner in nuclear material management and to provide references where detailed presentations can be found. Thus, this chapter provides a brief overview of some selected topics in simulation, but no attempt is made to introduce all relevant simulation techniques or to provide detailed discussions of all the necessary steps in a simulation study (e.g., variance reduction techniques to improve the efficiency of a simulation study are not discussed). Many excellent texts that provide detailed presentation and guidance have been written on simulation; among these are Bratley, Fox, and Schrage (1983); Fishman (1973); Kliejnen (1975); Law and Kelton (1982); and Rubinstein (1981).

11.1 THE ROLE OF SIMULATION

The philosophy behind developing and using system models in computerized simulation studies is that a simulator can be viewed as an artificial laboratory. That is, once a system is modeled and implemented on a computer, experiments can be performed using the model. These experiments, or simulations, permit inferences about a system (1) without building it, if it is only a proposed system; (2) without disturbing it, if it is an operating system that is costly or unsafe to experiment with; and (3) without destroying it, if the object of an experiment is to determine system limitations. In these ways, simulation studies are useful for design, procedural analysis, and performance assessment. Even when it is feasible to experiment directly with a system, experimenting with a computerized simulation of a system is often faster, easier, and/or cheaper than experimenting with the system. However, it must be emphasized that the validity of inferences made from the results of a simulation study depend on how closely the computerized model represents the operation of the system.

Simulation often has been viewed as a method of last resort, only to be employed when all other approaches fail. However, recent advances in simulation methodologies, availability of simulation software packages, and technical developments in electronics and computers have made simulation one of the most widely used tools in systems analysis, operations research, and the development and implementation of new statistical methods. Some advantages of simulation are:

1. Through simulation, it is possible to study the effects of certain informational, organizational, and environmental changes on the operation of a system by making alterations in the model of the system and then observing the effects of these alterations on the system behavior and response.
2. A detailed observation of the system being simulated may lead to a better understanding of the system and to suggestions for improving it, which otherwise would be unobtainable.
3. The experience of designing a computer simulation model is sometimes more valuable than the actual simulation itself, because the knowl-

edge gained in designing a simulation model frequently suggests changes in the system being simulated. The effects of these changes can then be tested via simulation before implementing them on the system.

4. Simulation of a complex system can yield insight into which variables are more important than others in the system and how these variables interact.

5. Simulation can be used to experiment with new situations about which there is little or no information in order to prepare for results of implementation on the system.

6. Simulation can serve as a preservice test to try out new policies and decision rules for operating a system, without running the risk of experimenting on the system.

7. When new elements are introduced into a system, simulation can be used to anticipate bottlenecks or other problems that could arise in the behavior of the system.

8. For certain types of stochastic problems (i.e., problems involving random variables), a simulated sequence of events may be of particular importance and may provide more valuable information than an analytic study involving only expected values.

9. Simulations can be performed to verify analytical solutions, to obtain approximate solutions when analytic solutions are difficult or impossible to obtain, and to try out new statistical methods to determine their properties and applicability in specific situations.

11.2 CONSIDERATIONS IN SIMULATION MODELING

In nuclear material research and management there is an emphasis on the development of symbolic or mathematical models to describe and analyze production, measurement, and quality control systems. These models range from pictorial maps of a production line to systems of differential equations describing chemical reactions. These models lend themselves to study by simulation because they can be manipulated rapidly and inexpensively on a computer to investigate the effects of changes in components of the system on the response or output of the system.

The development of a valid system model for use in simulation is often a major task and may require the expertise of mathematicians, statisticians, computer scientists, and other specialists, in collaboration with plant personnel who are knowledgeable about the operation and behavior of various aspects of the system. However, it is important for users to understand some of the considerations that go into model development so they can appreciate the uses and limitations of simulation. This section discusses some of the steps in developing and using a model. Specific details and illustrations are not provided, however, and the reader is encouraged to consult some of the texts listed in the references.

11.2.1 Types of Models

Simulation models are theoretically classified in two ways: as stochastic or deterministic models, and as discrete or continuous models. In practice, models are also classified by the way in which the form of the model is derived and the source of the values of the parameters. There are three types of models that are identified on the basis of these criteria. In this section, the theoretical characteristics are discussed first, and then the practical aspects are presented.

A **stochastic** model is one that includes randomness in the output so actual data are approximated. Randomness is introduced by incorporating random deviates in sections of the model where uncertainty occurs in the process being modeled. The topic of random deviate generation is introduced in Section 11.3. A **deterministic** model is one that is not stochastic.

To illustrate the concept of deterministic models, consider Newton's law of motion,

$$f = ma$$

This is a mathematical model that characterizes the motion of an object of mass m acted upon by a force f . This model is deterministic because, for example, the acceleration a produced by the force f is exactly determined by rearranging the equation to the form $a = f/m$. A more complicated deterministic model,

$$d = vt - 0.5gt^2$$

determines the distance d traveled in time t by an object thrown vertically with initial velocity v , where g is the acceleration due to gravity. Both of the above are deterministic models in the sense that they exactly determine the state of a physical process.

However, it is not always possible to derive a deterministic model for a physical process. Consider the simple process of tossing a coin. Theoretically, Newtonian mechanics could predict whether the coin will land with head or tail facing up (given initial velocity, spin, height, etc.), but in practice, not all factors that might affect how the coin lands can be specified in advance. Thus, a stochastic model of the process is constructed by assuming that each toss of the coin is a random trial which does not depend on the outcome of previous trials and that the coin is not weighted to favor its landing in a particular way. The resulting stochastic model for each toss of the coin is specified by

$$\text{Pr (Head)} = 0.5$$

$$\text{Pr (Tail)} = 0.5$$

For a sequence of n tosses of the coin, a stochastic model for the number of heads (X) is given by the binomial pdf (Equation 2.20), with $p = 0.5$. Although this stochastic model cannot be used to determine exactly what will happen in a single trial or in a sequence of trials, useful information can be gained about the physical process by studying such a model.

Discrete models occur when the model output can assume only a countable number of values at the time of each event. Models of many manufacturing processes are discrete because the output is measured by the number of finished items. However, the input to such a model does not have to be discrete. For example, nuclear fuel fabrication starts with a continuous flow of UF_6 powder and ends with discrete fuel rod assemblies. **Continuous** models are those in which the output changes continuously over time and can assume any value in some range at any given point in time. Models of the conversion process of UF_6 into UO_2 are continuous models.

In practice, almost all models are a combination of discrete and continuous parts. A model of a nuclear fuel fabrication facility is an example of such a combined model. Input to the plant arrives as cylinders of UF_6 powder and components for fuel rod assemblies. The inventory of these items is modeled with a discrete system. When the UF_6 cylinders are removed from inventory, their contents are fed into a continuous chemical process that produces UO_2 . A continuous system model of the chemical process can then be used to predict output rate or to find the conditions which optimize yield. The blending of powders for fuel pellets is also modeled continuously. The pressing, sintering, inspecting, and storing of fuel pellets are discrete, but they are modeled as a continuous system because the number of pellets is too large to be directly counted in practice. The options are to model total weight or number of pellets with a continuous model, or to model the number of trays or racks of pellets with a discrete model. Fuel rod assembly, inspection, and inventory are modeled with a discrete system because the items are countable. Scrap inventory requires a combination of discrete and continuous models; barrels and drums of scrap powder, pellets, and liquids are modeled continuously by weight or discretely by the number of containers. Inventory of defective fuel rods is modeled discretely. Finally, recovery of uranium from scrap is modeled continuously.

Some components of a model of a fuel fabrication facility are stochastic and some are deterministic. Inventory models of UF_6 cylinders or fuel assemblies are deterministic because these items can be counted, and it is assumed there are no counting errors. (In practice, accurate counting of fuel rods is a large task which may delay reconciliation of inventory by hours or even days). The amount and uranium content of UF_6 powder in a cylinder and the input to the chemical system that produces UO_2 are stochastic because there may be errors in the chemical analyses and weigh-

ings. The chemical system is modeled stochastically; the randomness comes from small variations in such variables as temperature, flow, and the condition of the machinery. Modeling a chemical system is extremely complex. It involves many factors—some not identified, some not understood, some not controllable—with complex interrelationships. The pellet production model, whether continuous or discrete, is stochastic because the number of pellets is too large to count accurately, and the weighing of trays or buckets of pellets is subject to error. Models of the number of barrels of scrap on inventory are deterministic, but models of scrap content are stochastic because the heterogeneous nature of scrap does not allow accurate assay of drum content.

The methods used to determine the form of the model and to obtain values of the model parameters provide another way to classify models. These two considerations give rise to the following three types of models:

1. The form of the model is known from theory and the parameters have known values. The theory on which the models are based does not have to be unequivocally established and accepted. The parameters may not be theoretical values but may have been established to a high degree of accuracy through experimentation, so the parameters are treated as known values. This is the most reliable type of model for simulation use.

2. The form of the model is known based on theory, but the parameters are estimated from experimental data. This type of situation is more common than that given in (1).

3. The form of the model is developed to approximate reality; the model is empirically determined and is highly dependent in form on the data available to create the model. The parameters may be estimated from the same data used to determine the form of the model. This is the least reliable type of model for simulation use, but it probably is the type used most frequently.

Whether a model is of form (1), (2), or (3), it can be used to some advantage in a simulation study. Many instances involve predicting a response or several responses corresponding to a given set of prescribed conditions or estimating the uncertainty associated with a response or several responses. The response is calculated from the model, and the uncertainty is calculated by either error propagation techniques or by simulation methods, depending on the complexity of the model.

Because the validity of results obtained from using a model depends on the quality of the model, methods of model development are important. Not every attempt to model a process is completely successful. However, the success in specifying the model form and in estimating its parameters determines the quality of a model. A good quality model is the first step toward answering research questions using simulation.

11.2.2 Development of a Model and Study Plan

The first developmental activity is to design and implement the simulation model. The specific tasks in this activity are not discussed in detail in this section. However, excellent discussions are found in many simulation textbooks, such as Bratley, Fox, and Schrange (1983); Emshoff and Sisson (1970); Meier, Newell, and Pazer (1969); and Naylor et al. (1966). Assuming that these tasks are completed, the next activity is model verification. A verified model is one that has been shown to represent the system or process as the designer intended. Verification is an important activity. Without satisfactory and explicit verification, it is possible to have a model that appears to work satisfactorily but gives answers that are erroneous. Bratley, Fox, and Schrange (1983) and Fishman and Kiviat (1967) suggest techniques for verification, including statistical methods, that go far beyond the practices of simple comparative analysis and manual checking of calculations.

Bratley, Fox, and Schrange (1983) and Fishman and Kiviat (1967) also describe model validation methods. Some additional suggestions on validation techniques are provided by Naylor (1971). Validation is one of the most critical activities performed in any simulation study. It is also one of the most difficult to accomplish satisfactorily. A validated model is one that has been shown to be a reasonable abstraction of the system it is intended to represent. The usual approach to validation is to run the model with historical data and compare model results with actual system results for the same historical period. However, such comparisons may not be possible if the model is experimental or predictive in nature. It also may be difficult to make a statistical comparison of results when such a comparison is appropriate because of the requirement that equilibrium be reached before results are measured. It may take considerable time to reach equilibrium with the computer simulation model, while the real-world system might never exist in a state of equilibrium, thereby severely complicating the comparison.

The next developmental activity is designing and planning the experimentation to be done with the simulation model. This includes specifying the information to be determined and the desired accuracy of that information. The two classes of experimentation are exploration of system behavior and optimization of system parameters. Exploration of system behavior is undertaken in an attempt to explain the relationship between results and the controllable parameters of the simulation. Optimization is performed to find the combination(s) of parameter levels that minimize or maximize the results of the simulation. Experimental designs like full factorials, fractional factorials (Chapter 7), or special-purpose response surface designs—such as those presented by Myers (1976)—are appropriate

for exploratory experimentation. For optimization, optimum-seeking techniques are available, and though in some cases they cannot guarantee a global optimum, they can provide very useful results. Hunter and Naylor (1970) give a good outline of experimental designs for simulation studies, and Schmidt and Taylor (1970) have a good discussion on simulation optimization techniques. Bohachevsky, Johnson, and Stein (1986) introduce a generalized simulated annealing method for optimization of functions (possibly constrained) that have many local extrema.

11.2.3 Operational Activities

At the stage in the simulation process when the model has been designed, implemented, and studied, and its use has been planned, the remaining step is to carry out the actual simulation experiment. This must include tactical design of the experiments to be performed, such as determining how many simulation runs to execute and how to collect data from each run. This involves establishing initial conditions for model variables and estimating parameters so the simulated system will reach a state of equilibrium as soon as possible. Criteria for recognizing equilibrium must also be determined so data can be gathered without being biased by transients from the startup of the run. Other considerations are the sample size required for data collection and the techniques to be used to compare alternative systems (if this is the study objective). In the latter case, relative results from the simulation runs will be of interest. Such methods as using the same sequence of random numbers for each run can be applied; this will reduce the residual variation between sets of results and thus reduce the length of the simulation runs themselves.

In the simulation runs the model is run for the specified time, the parameters are changed, and the model is run again. This process is repeated until the experimentation is completed and is followed by an analysis of the simulated data. Techniques such as regression analysis (Chapter 6) and analysis of variance (Chapter 5) are widely used for interpreting simulation data, depending on the original objectives.

If the objectives have been met, the simulation study is completed at this point. Because simulation is often a trial-and-error process, however, the objectives may not have been satisfied, which leaves two general alternatives available. The first is to modify the model to facilitate discrimination among simulated systems and then rerun the experiment. The second alternative is to use the original model but alter the design of the experiment, using new search techniques or more extensive experimental designs.

11.3 GENERATION OF RANDOM DEVIATES

Stochastic simulations require methods of generating random deviates from predetermined distributions. The most common methods require gen-

erating random deviates from the uniform distribution (Section 2.5.1) and then transforming these to deviates from the desired distribution. Independent random deviates that are uniformly distributed on the interval 0 to 1 inclusive, denoted by $[0,1]$, are called **random numbers** and provide the basis for generating random deviates from other distributions. In this section, methods of generating random numbers are discussed and the techniques used to obtain random deviates from a few commonly used distributions are given. Some numerical examples are provided to illustrate how random deviates are used to solve specific problems by simulation analysis. Most of the example problems are easily solved by methods presented in previous chapters. They are used here for simplicity of illustration, however, to compare simulation results with analytic results. It is hoped this demonstrates the utility and power of simulation as a tool for solving more complex problems.

Note that the methods presented in this section for generating random deviates are not necessarily the most currently developed or the best techniques available. This brief presentation is intended to lend some intuitive understanding of the subject of generating random deviates. Only a few probability distributions are addressed. For further guidance, the reader is encouraged to consult some of the referenced texts, especially the more recent ones [e.g., Bratley, Fox, and Schrage (1983) or Law and Kelton (1982)].

11.3.1 Random Numbers

No numerical algorithm generates a truly random number; for further discussion of this concept, see Kolata (1986). However, hardware-dependent algorithms are available that produce numbers close enough to randomness to be used in simulation. Before a sample of random numbers is used, it is important to ascertain that (1) the random numbers are distributed uniformly; (2) the random sample, in the order in which it is generated, does not contain serial correlations; and (3) the runs in the random sample are of random length and number. If multiple random samples are generated for use, it is also important to ensure that the samples are independent. An overview of problems associated with random number generation is given by Marsaglia (1968), and methods of generating and testing random numbers are presented in Kennedy and Gentle (1980).

A random number generating algorithm widely used on computers is the congruential generator, symbolically given by:

$$z_{i+1} = (az_i + b)(\text{mod } c), \quad i = 0, 1, 2, 3, \dots \quad (11.1)$$

$$r_{i+1} = z_{i+1}/c$$

Here z_0 is a seed value provided by the user and r_i is the i^{th} random number, standardized to be in the interval $[0,1]$. The seed value, z_0 , is chosen to be a large odd positive integer within the constraints of the hardware being used. It is advantageous to be able to repeat a simulation exactly; thus, the value of z_0 always should be recorded. Equation 11.1 denotes that the unstandardized random number z_{i+1} is equal to the remainder of $(az_i + b)$ divided by c , where z_i is the previous unstandardized random number, and where a , b , and c are constants. The period of a generator is the number of random numbers generated before the series repeats itself. The constants a , b , and c are chosen so that the period is as long as possible while retaining the independence of the numbers. Typically, $c = 2^B$, where B is the number of bits per word for the computer being used. The constant b is chosen to be a small, positive integer relatively prime to c , so that the greatest common factor of b and c is unity. The constant a is chosen to be $a = 1 + 4k$ where k is a positive integer.

Before the availability of computers, tables were used to generate random numbers. Although tables of random numbers are not widely used now, they are discussed here for completeness. Tables of random numbers are generated so that all digits (0 through 9) appear with equal probability at any position in the table, and so that no recognizable patterns (serial correlations and run length or number) exist in prescribed sized groups of random numbers. A well-known table of random numbers was published by the Rand Corporation (1955). While random number tables are useful for randomization of a small sampling plan or experimental design, they should not be seriously considered for use as the random number generator in a simulation study of any size.

To use a table of random numbers, a random starting place is selected. This is done by arbitrarily pointing to a location on an arbitrary page and taking the first appropriate digits as the starting page, row, and column. If a table has 50 pages, the first 2 digits >00 and ≤ 50 determine the starting page. If a table has 100 rows and 100 columns, each numbered 0 to 99, the next 2 digits ≤ 99 determine the starting row, and the succeeding 2 digits ≤ 99 determine the starting column. In this manner, the starting place in the table is located for generating one sample of random numbers. The sample of random numbers is then obtained by following the column(s) down the rows of the table and recording the numbers in the order they appear.

After a population of objects is numbered from 1 to N , a random number table is used to select a random sample of size n by letting n digits selected from the table identify the objects to be in the sample. If a combination of digits does not correspond to an object in the population, the combination is discarded. For example, suppose a random sample of size $n = 10$ from a population of size $N = 46$ is required. Two-digit random numbers are selected, ignoring any combinations >46 or equal to 00. If

sampling is with replacement, the first 10 random numbers within the acceptable range selected from the table specify the sample. If sampling is without replacement, each acceptable number is used only the first time it is encountered. Thus, in the previous example, if sampling is without replacement and the combination of digits 06 is encountered twice, it is used only the first time. A total of 11 random numbers (not counting other unallowable pairs of digits that are >46 or equal to 00) are then selected so that the sample size is $n = 10$. Finally, the observed random numbers determine which objects from the population are to be included in the sample. For example, if the first 2 random numbers are 06 and 25, then the 6th and 25th objects are 2 of the 10 objects included in the sample. A more computer-oriented discussion of this application appears in Section 2.5.1.

To generate random numbers on a hand-held calculator, the directions accompanying the calculator must be followed. For most hand-held calculators, generating distinct multiple samples of a given size requires providing a new seed value for each sample. If a new seed value is supplied, a random sample should be tested to assure that the random numbers are uniformly distributed, that they are not serially correlated, and that the lengths and numbers of runs are random. Tests of independence should also be used to assure that multiple random samples are independent. The reader is cautioned that at the time this book was being prepared, there was considerable debate over the randomness and validity of random number generators on hand-held calculators.

11.3.2 The Uniform Distribution

The uniform distribution is described in Section 2.5.1. Random numbers with multiple digits and a decimal point in front of the leading digit are approximately uniformly distributed with parameters $\alpha = 0$ and $\beta = 1$. Such random numbers are said to be random deviates from the **unit uniform distribution**. Note that the uniform distribution is defined as a continuous distribution, but random numbers are actually discrete. Thus, random numbers with a decimal point in front of the leading digit are only approximately uniformly distributed. A random deviate x from a uniform distribution with parameters $\alpha = a$ and $\beta = b$, where $\alpha < \beta$, is obtained from a unit uniform random deviate r by the transformation

$$x = r(b-a) + a \quad (11.2)$$

For example, if the 5 digit random number 34974 is chosen randomly from a table of random numbers, then the number 0.34974 is a random deviate from the unit uniform distribution. A random deviate x from the uniform distribution with $\alpha = 9$ and $\beta = 11$ is obtained by applying Equation 11.2, which gives $x = (0.34974)(11-9) + 9 = 9.69948$.

The following example illustrates one of the many practical problems that can be solved via simulation using uniform random deviates.

Example 11.1 Consider the problem of evaluating a definite integral

$$\int_a^b f(x)dx$$

where $f(x)$ may be difficult to integrate. To compare the example results with analytic results, consider evaluating the probability that an observed value of the random variable X from the standard normal distribution will fall in the interval $-1.3 \leq X \leq 1.6$. The desired probability is obtained by evaluating the integral

$$\int_{-1.3}^{1.6} \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-x^2}{2}\right] dx$$

Of course, this integral can be directly evaluated analytically or Table A3 can be used. However, the following solution illustrates how simulation can be used to approximate the value of integrals that are difficult to work with analytically.

The first step is to construct a rectangle that “frames” the function in the region of the integration. The base of the rectangle is a horizontal line from a to b , so that the width of the rectangle is defined by the limits of integration to be

$$W = b - a$$

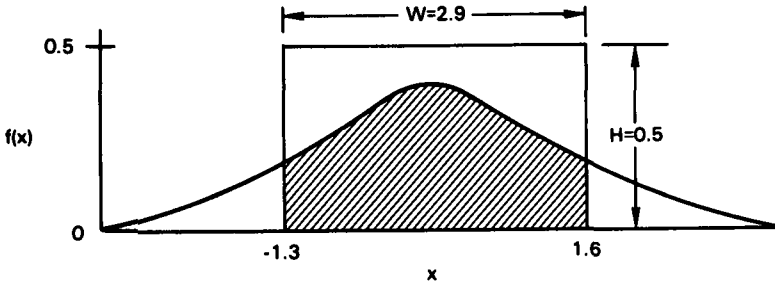
The height H of the rectangle must be chosen so that the function $f(x)$ is inside the rectangle over the range $a \leq x \leq b$. The area of the rectangle is then

$$A = WH = (b - a)H$$

For the standard normal example, the width of the rectangle (pictured below) is $W = 1.6 - (-1.3) = 2.9$, and the base of the rectangle is a line along the horizontal axis connecting the points $x = -1.3$ and $x = 1.6$. The maximum value of $f(x)$ on the range $-1.3 \leq x \leq 1.6$ occurs at $x = 0$, where $f(x) = 1/\sqrt{2\pi} = 0.3989$. Thus, the height of the framing rectangle must exceed 0.3989; for this example, $H = 0.5$ is chosen. The

area of the rectangle is $A = 2.9(0.5) = 1.45$. In the rectangle pictured below, the shaded region is the probability to be evaluated.

The next step is to generate a large number, say N , of pairs of uniform random deviates (x,y) , where x is from the uniform distribution over the interval $[-1.3,1.6]$ and y is from the uniform distribution over the interval $[0,0.5]$. Each pair (x,y) is obtained by generating two independent unit



uniform random deviates r_1 and r_2 , and then using Equation 11.2 to compute

$$x = r_1[1.6 - (-1.3)] + (-1.3) = 2.9r_1 - 1.3$$

and

$$y = r_2(0.5 - 0.0) + 0.0 = 0.5r_2$$

The function $f(x)$ is then evaluated for each observed value of x . Let n denote the number of pairs where $y \leq f(x)$. Then the value of the definite integral is approximated by

$$\frac{nA}{N} = \frac{n(b-a)H}{N} = \frac{n(1.45)}{N}$$

For this example, $N = 1000$ pairs (x,y) were generated on a computer using the method described previously. Results for the first five pairs are tabulated below for illustration. Of the $N = 1000$ pairs, there are $n = 601$ pairs where $y \leq f(x)$. Thus, the approximate value of the integral of interest is

$$\frac{601(1.45)}{1000} = 0.8715$$

Note that the actual value of the integral, obtained by using Table A3, is 0.8484. For many applications, the accuracy of this approximation

| r_1 | r_2 | x | y | $f(x)$ |
|--------|--------|---------|--------|---------|
| 0.5825 | 0.4014 | 0.3893 | 0.2007 | 0.3698† |
| 0.7950 | 0.5618 | 1.0056 | 0.2809 | 0.2406 |
| 0.5540 | 0.5484 | 0.3068 | 0.2742 | 0.3806† |
| 0.1593 | 0.2563 | -0.8380 | 0.1281 | 0.2808† |
| 0.6839 | 0.0436 | 0.6833 | 0.0218 | 0.3159† |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

† $y \leq f(x)$

would be satisfactory. However, a more accurate approximation can be obtained by increasing N .

A second simulation method, that may be more accurate for some functions than the method described previously, is based on the approximation

$$\int_a^b f(x)dx \cong \frac{(b-a)}{N} \sum_{i=1}^N f(x_i)$$

The method involves approximating the integral by generating a large number N of uniform random deviates x from the interval $[a,b]$, where the N deviates are substituted for the x_i in this formula.

For the standard normal example, $N = 1000$ uniform deviates x were generated from the interval $[-1.3, 1.6]$ as discussed previously. Then the function $f(x)$ was evaluated at each of the x values, and the $N = 1000$ $f(x)$ values were summed, giving $\sum f(x) = 290.45$. The approximate value of the integral of interest is then

$$\frac{(b-a)}{N} \sum_{i=1}^N f(x) = \frac{2.9}{1000} (290.45) = 0.8423$$

Note that this is much closer to the actual value 0.8484 than the approximation obtained by the rectangular method, and increasing N will improve the accuracy of this approximation.

11.3.3 The Normal Distribution

The normal distribution is described in Section 2.5.2. A method for generating random deviates from the normal distribution was first intro-

duced by Box and Muller (1958). The Box-Muller technique is a method still in wide use, even though much faster algorithms are currently available. The mathematical development of this method is not presented here, only the results necessary for its implementation. For more details, see Bratley, Fox, and Schrage (1983), Law and Kelton (1982), or Kennedy and Gentle (1980).

Two independent random deviates, z_1 and z_2 , from the standard normal distribution are obtained by generating two unit uniform random deviates, r_1 and r_2 , and then letting

$$z_1 = [2\ln(1/r_1)]^{1/2} \cos(2\pi r_2)$$

and (11.3)

$$z_2 = [2\ln(1/r_1)]^{1/2} \sin(2\pi r_2)$$

Because z_1 and z_2 are independent random deviates from the normal distribution with mean 0 and standard deviation 1, they can be transformed into two independent random deviates, x_1 and x_2 , from a normal distribution with mean μ and standard deviation σ by the transformation

$$x = \sigma z + \mu \tag{11.4}$$

An improvement on the Box-Muller method, which eliminates the trigonometric calculations, was introduced by Marsaglia and Bray (1964). It has become known as the polar method, and was found by Atkinson and Pearce (1976) to be as much as 31% faster than the Box-Muller method, depending on which computer is used. More recently, a very fast algorithm for generating random deviates from the normal distribution was developed by Kinderman and Ramage (1976). Although somewhat more complicated, this algorithm proved to be 30% faster than the polar method in a series of experimental comparisons. The details of these newer methods are not presented here, and the Box-Muller method is used throughout the remaining sections of this chapter to generate random deviates from the normal distribution.

Example 11.2 To illustrate the Box-Muller technique, suppose that it is of interest to generate a random sample of size $n = 20$ from a normal distribution with mean $\mu = 20$ and standard deviation $\sigma = 0.5$. The first step is to generate 10 pairs of independent unit uniform random deviates (r_1, r_2) . Then Equations 11.3 and 11.4 are applied with $\mu = 20$ and $\sigma =$

0.5. These steps were performed on a computer and the results are summarized below.

| r_1 | r_2 | z_1 | z_2 | x_1 | x_2 |
|----------|----------|----------|----------|---------|---------|
| 0.143624 | 0.154317 | 1.11432 | 1.62463 | 20.5572 | 20.8123 |
| 0.961072 | 0.528347 | -0.27734 | -0.04093 | 19.8613 | 19.9750 |
| 0.285263 | 0.423452 | -1.40419 | 0.73276 | 19.2979 | 20.3664 |
| 0.833866 | 0.405667 | -0.49998 | 0.33673 | 19.7500 | 20.1684 |
| 0.268445 | 0.982856 | 1.61239 | -0.17436 | 20.8062 | 19.9128 |
| 0.239110 | 0.896121 | 1.34393 | -1.02738 | 20.6720 | 19.4863 |
| 0.110780 | 0.206207 | 0.56995 | 2.01881 | 20.2850 | 21.0094 |
| 0.460185 | 0.487190 | -1.24186 | 0.10017 | 19.3791 | 20.0501 |
| 0.503274 | 0.710657 | -0.28674 | -1.13623 | 19.8566 | 19.4319 |
| 0.625828 | 0.353656 | -0.58692 | 0.77000 | 19.7065 | 20.3850 |

The 20 values in the last two columns are the equivalent of a random sample of size $n = 20$ from a normal distribution with mean $\mu = 20$ and standard deviation $\sigma = 0.5$.

Another method for generating random normal deviates, which is much less efficient than the methods discussed previously, is based on the Central Limit Theorem (Section 2.5.2.1) which states that the sum of a large number of variables from any distribution is approximately normally distributed. If the expected value of such a sum is subtracted from the observed sum and this difference is divided by the standard deviation of the sum, a deviate with approximately the standard normal distribution is obtained.

To apply this method, generate a large number, say k , of unit uniform random deviates r_1 through r_k . When k is sufficiently large, the quantity z , given by

$$z = \frac{\sum_{i=1}^k r_i - k/2}{\sqrt{k/12}} \quad (11.5)$$

approximates a random deviate from the standard normal distribution. According to the rules of thumb presented in Section 2.5.2.1 for applying the Central Limit Theorem, $k = 12$ is adequate, and Equation 11.5 reduces to the form

$$z = \sum_{i=1}^{12} r_i - 6$$

A random deviate x from a normal distribution with mean μ and standard deviation σ is then obtained by applying Equation 11.4. This method is not recommended for general use and is presented only for information and illustrative purposes.

Example 11.3 To illustrate the method based on the Central Limit Theorem, suppose that it is of interest to generate a random sample of size $n = 3$ from a normal distribution with mean $\mu = 20$ and standard deviation $\sigma = 0.5$. First, 3 random samples of 12 unit uniform deviates are generated. Then, to each sample, Equation 11.5 is applied with $k = 12$, and Equation 11.4 is applied with $\mu = 20$ and $\sigma = 0.5$. These steps were carried out on a computer with the following results.

| | Sample 1 | Sample 2 | Sample 3 |
|----------------|----------|----------|----------|
| | 0.993458 | 0.873662 | 0.742130 |
| | 0.172875 | 0.986846 | 0.203562 |
| | 0.339441 | 0.486923 | 0.857184 |
| | 0.834880 | 0.282859 | 0.822082 |
| | 0.296872 | 0.790098 | 0.400560 |
| | 0.624603 | 0.291086 | 0.248743 |
| | 0.700297 | 0.033209 | 0.453487 |
| | 0.844989 | 0.716828 | 0.877085 |
| | 0.517123 | 0.619234 | 0.372241 |
| | 0.170803 | 0.899237 | 0.326297 |
| | 0.162543 | 0.392950 | 0.036732 |
| | 0.677929 | 0.655703 | 0.072874 |
| Σr_i : | 6.3358 | 7.0286 | 5.4130 |
| z : | 0.3358 | 1.0286 | -0.5870 |
| x : | 20.1679 | 20.5143 | 19.7065 |

The last row, labeled x , is a random sample of size $n = 3$ from a normal distribution with mean $\mu = 20$ and standard deviation $\sigma = 0.5$. Note that for about the same effort, the Box-Muller method would have produced 36 random normal deviates.

Although simulation can be used to study quite complicated problems, the following relatively simple example illustrates one of the basic types of

problems where a simulation analysis can provide useful information. The example problem is solved analytically in Chapter 4.

Example 11.4 In Example 4.3, a sampling plan is developed for monitoring the percent uranium in batches of UO_2 powder. For a given batch, the plan is to select randomly $n = 7$ powder samples and determine the percent uranium in each sample. Then, if the computed value of the test statistic

$$Z = \frac{\bar{X} - 87.6}{0.03/\sqrt{7}}$$

is less than or equal to -1.65 , the null hypothesis $H_0: \mu \geq 87.6$ is rejected at the 0.05 level of significance in favor of the alternative hypothesis $H_A: \mu = \mu_A < 87.6$. The plan was developed to meet the requirement that the power of the test must be at least 0.90 whenever the true mean is as small as $\mu_A = 87.565$.

The points for constructing a power curve for this plan are determined analytically and tabulated in Example 4.3. Suppose that it is of interest to verify these power calculations by simulation analysis.

When the value of the mean is μ_A , the test statistic Z follows a normal distribution with mean $(\mu_A - 87.6)/(0.03/\sqrt{7})$ and standard deviation 1. Thus, a simulation analysis to approximate the power of the test when the mean is μ_A is carried out in the following steps:

1. Set the value of μ_A , and generate N random deviates from the normal distribution with mean $(\mu_A - 87.6)/(0.03/\sqrt{7})$ and standard deviation 1.
2. Count the number of deviates, say n , which are less than or equal to -1.65 .
3. The approximate power of the test when the mean is μ_A is $1 - \beta \cong n/N$. The accuracy of this approximation improves as N increases. Thus, a large value of N should be used.
4. Repeat steps 1, 2, and 3 for each value of μ_A to be studied.

For this example N was chosen to be 10,000, and the Box-Muller technique was used on a computer. The results are summarized below. Note the close agreement between the probabilities in the $1 - \beta$ column above the corresponding probabilities obtained analytically in Example 4.3.

| μ_A | n | $n/N \cong 1 - \beta$ |
|---------|------|-----------------------|
| 87.550 | 9968 | 0.9968 |
| 87.555 | 9904 | 0.9904 |
| 87.560 | 9705 | 0.9705 |
| 87.565 | 9257 | 0.9257 |
| 87.570 | 8404 | 0.8404 |
| 87.575 | 7074 | 0.7074 |
| 87.580 | 5419 | 0.5419 |
| 87.590 | 2243 | 0.2243 |
| 87.600 | 495 | 0.0495 |
| 87.800 | 65 | 0.0065 |

11.3.4 The Chi-Square Distribution

The chi-square distribution, discussed in Sections 2.5.4 and 3.5.1, is the underlying distribution of the sample variance computed from the observations in a random sample from a normal distribution. Thus, in simulation modeling and analyses, it is often necessary to generate random deviates from the chi-square distribution with ν degrees of freedom.

One approach is to generate ν random deviates z_i from the standard normal distribution and compute the quantity

$$x = \sum_{i=1}^{\nu} z_i^2 \quad (11.6)$$

Then x is a random deviate from the chi-square distribution with ν degrees of freedom.

Although this method yields satisfactory results, it is much slower for large ν than necessary. A quicker method makes use of the fact that the chi-square distribution with ν degrees of freedom is a gamma distribution (Table 2.5, Section 2.5.4) with $\lambda = 1/2$ and $r = \nu/2$. Then x can be obtained directly from the gamma generation methods presented in most texts on simulation or random number generation. See, for example, Bratley, Fox, and Schrage (1983); Law and Kelton (1982); Fishman (1973); or many of the other texts listed in the references.

When a random sample of size n is drawn from a normal distribution with unknown mean μ and unknown standard deviation σ , the quantity

$$x = \frac{(n-1)S^2}{\sigma^2} \quad (11.7)$$

is an observation from the chi-square distribution with $n-1$ degrees of freedom, where S^2 is the sample variance. Thus, to generate a sample variance S^2 for a specified sample size n and specified value of σ^2 , generate a random deviate x from the chi-square distribution with $n-1$ degrees of freedom. Then substitute x , σ^2 , and $n-1$ into Equation 11.7 and solve for S^2 . This gives

$$S^2 = \frac{x \sigma^2}{n-1}$$

There is a simpler method for generating S^2 for a sample of size n from the $N(\mu, \sigma^2)$ distribution. First generate n random normal deviates from the $N(\mu, \sigma^2)$ distribution by applying the Box-Muller or an equivalent method. Then compute S^2 directly. Note, however, that this method requires more calculations than the method presented above, and thus is slightly slower.

Example 11.5 In Example 4.9, a sampling plan is defined that ensures a power of at least 0.80 when the true value of σ^2 is twice as large as the hypothesized value $\sigma_0^2 = (0.2 \text{ kg})^2$. The plan requires that a random sample of size $n = 20$ containers be drawn and individually weighed. The test statistic

$$\chi^2 = \frac{(n-1)S^2}{(0.2)^2} = \frac{19S^2}{(0.2)^2}$$

is computed, where S^2 is the sample variance computed from the observed container weights. If $\chi^2 \geq 27.2$, the null hypothesis $H_0: \sigma \leq 0.2 \text{ kg}$ is rejected at the 0.10 level of significance in favor of the alternative hypothesis $H_A: \sigma = \sigma_A > 0.2 \text{ kg}$. It is of interest to verify that the power of the hypothesis test is at least 0.80 when $\sigma_A^2/(0.2)^2 = 2$.

When H_A is true, the quantity

$$\frac{(n-1)S^2}{\sigma_A^2}$$

is an observation from the chi-square distribution with $n-1$ degrees of freedom. Note that the test statistic χ^2 is obtained by multiplying the above quantity by $\sigma_A^2/(0.2)^2$. That is,

$$\chi^2 = \frac{(n-1)S^2}{(0.2)^2} = \left[\frac{\sigma_A^2}{(0.2)^2} \right] \frac{(n-1)S^2}{\sigma_A^2}$$

This means that when H_A is true, the test statistic χ^2 is equal to an observation from the chi-square distribution multiplied by the quantity $\sigma_A^2/(0.2)^2$.

In this example, $\sigma_A^2/(0.2)^2 = 2$. Thus, the power of the test is determined from simulation analysis by completing the following steps:

1. Generate $n-1 = 19$ random deviates, z_1 through z_{19} , from the standard normal distribution, and compute a chi-square deviate x using Equation 11.6. Then compute a new deviate $y = (\sigma_A^2/\sigma_0^2)x = 2x$.
2. Step 1 is repeated M times, where M is chosen to be quite large.
3. Count the number, say m , of deviates y that are greater than or equal to 27.2. The approximate power of the test is $1-\beta \cong m/M$.

These steps were carried out on a computer with $M = 5000$. There were $m = 4021$ deviates y that were greater than or equal to 27.2. Thus, the power of the test when $\sigma_A^2/(0.2)^2 = 2$ is approximately

$$1-\beta \cong 4021/5000 = 0.8042$$

This verifies that the proposed plan meets the specified power requirement.

11.3.5 The Central and Noncentral Student's t Distributions

Explicit forms of the central and noncentral Student's t distributions are not given in any of the previous chapters, even though they play an important role in statistical inference. A random variable X is said to have the (central) Student's t distribution with ν degrees of freedom if X can be expressed as the ratio of a standard normal random variable to the square root of a chi-square random variable with ν degrees of freedom divided by its degrees of freedom (Equation 3.25). This definition provides a method for generating a random deviate from the (central) Student's t distribution with ν degrees of freedom.

The procedure is to generate $\nu + 1$ random deviates z_i from the standard normal distribution and compute the quantity

$$x = \frac{z_1}{\left[\sum_{i=2}^{\nu+1} z_i^2 / \nu \right]^{1/2}} \quad (11.8)$$

Then x is a random deviate from the Student's t distribution with ν degrees of freedom.

To illustrate this method, suppose that it is of interest to generate a random deviate from a Student's t distribution with $\nu = 9$ degrees of freedom. In the table of numbers in Example 11.2, the column headed z_1 includes 10 random deviates from the standard normal distribution. If these deviates are labeled z_1 through z_{10} , then applying Equation 11.8 gives

$$x = \frac{z_1}{\left(\sum_{i=2}^{10} z_i^2 / 9 \right)^{1/2}}$$

$$= 1.11432 / \sqrt{8.99835/9} = 1.11442$$

which is a random deviate from the Student's t distribution with $\nu = 9$ degrees of freedom.

To generate a random deviate x from the noncentral Student's t distribution with ν degrees of freedom and noncentrality parameter λ , Equation 11.8 is modified to the form

$$x = \frac{z_1 + \lambda}{\left(\sum_{i=2}^{\nu+1} z_i^2 / \nu \right)^{1/2}} \quad (11.9)$$

For example, to generate a random deviate from the noncentral Student's t distribution with $\nu = 9$ degrees of freedom and noncentrality parameter $\lambda = 1.0$, the results from the previous example are used in Equation 11.9 to compute

$$x = \frac{1.11432 + 1.0}{\sqrt{8.99835/9}} = 2.11451$$

Example 11.6 In Example 4.6, a sampling plan is developed based on an approximate sample size formula with a sample estimate substituted for the unknown standard deviation. The objective is to determine the sample size n required to ensure that the power of the hypothesis test is at least 0.90 when the true mean percent uranium in a batch of UO_2 powder

is as small as $\mu_A = 87.565$. The plan developed in Example 4.6 is to randomly select $n = 9$ powder samples from a given batch, determine the percent uranium in each sample, and compute the test statistic

$$t = \frac{\bar{X} - 87.6}{S/\sqrt{9}}$$

If $t \leq -1.860$, the null hypothesis $H_0: \mu \geq 87.6$ is rejected at the 0.05 level of significance in favor of the alternative hypothesis $H_A: \mu = \mu_A < 87.6$.

It is assumed that the observed percent uranium values for the $n = 9$ powder samples are a random sample from a normal distribution with unknown mean μ and unknown standard deviation σ . If H_0 is true, the test statistic t is an observation from the central Student's t distribution with $\nu = 8$ degrees of freedom. If H_A is true, t is an observation from the non-central Student's t distribution with $\nu = 8$ degrees of freedom and with noncentrality parameter $\lambda = (\mu_A - \mu_0)/(\sigma/\sqrt{n}) = 3(\mu_A - 87.6)/\sigma$.

Although σ is unknown, one goal of the process manager is to control σ so that $\sigma \leq 0.03$. Although this level of control has not yet been established, it is believed that σ will not exceed 0.033 for any batch of UO_2 powder.

It is of interest to investigate the power of the hypothesis test for the mean percent uranium under the proposed plan when $\mu_A = 87.565$, for values of σ ranging from 0.030 to 0.033. This can be accomplished using simulation analysis, by generating random deviates from the noncentral Student's t distribution as follows:

Consider only the values $\sigma = 0.030, 0.031, 0.032, \text{ and } 0.033$. For each value of σ , perform the following steps:

1. Generate nine random deviates, z_1 through z_9 , from the standard normal distribution. Set $\lambda = 3(87.565 - 87.6)/\sigma = -0.105/\sigma$, and use Equation 11.9 to compute the random deviate

$$x = \frac{z_1 + \lambda}{\left(\sum_{i=2}^9 z_i^2 / 8 \right)^{1/2}}$$

2. Step 1 is repeated M times, where M is chosen to be large.
3. Count the number, say m , of deviates x which are less than or equal to -1.860 . The power of the test, when $\mu_A = 87.565$ for the specified σ , is then approximately $1 - \beta \cong m/M$.

These steps were carried out on a computer with $M = 10,000$ for each of the four values of σ . The results are tabulated as follows:

| σ | m | $1 - \beta \cong m/10,000$ |
|----------|------|----------------------------|
| 0.030 | 9361 | 0.9361 |
| 0.031 | 9218 | 0.9218 |
| 0.032 | 9064 | 0.9064 |
| 0.033 | 8894 | 0.8894 |

These results indicate that the power of the test is greater than or equal to 0.8894 when the true mean percent uranium in a batch is $\mu_A = 87.6$ and $\sigma \leq 0.033$. Strictly speaking, the proposed plan is adequate (i.e., $1 - \beta \geq 0.90$) only if $\sigma \leq 0.032$. However, for practical purposes, a power of 0.8894 may be satisfactorily close to 0.90; if so, the proposed plan is satisfactory for $\sigma \leq 0.033$.

11.3.6 The F Distribution

If x_1 and x_2 are independent random deviates from chi-square distributions with degrees of freedom ν_1 and ν_2 , respectively, then

$$x = \frac{x_1/\nu_1}{x_2/\nu_2} = \frac{\nu_2 x_1}{\nu_1 x_2} \quad (11.10)$$

is a random deviate from the F distribution with ν_1 and ν_2 degrees of freedom. Thus, to generate a random deviate x from the F distribution, x_1 and x_2 are generated as described in Section 11.3.4 and are substituted into Equation 11.10.

Example 11.7 In Example 4.14, a sampling plan is developed for testing the hypothesis

$$H_0: \sigma_1^2/\sigma_2^2 \leq 1$$

against the alternative

$$H_A: \sigma_1^2/\sigma_2^2 > 1$$

at the 0.10 level of significance. The plan requires taking independent ran-

dom samples of size $n_1 = 14$ measurements using the rapid method and $n_2 = 7$ measurements using the standard method. The test statistic

$$F = S_1^2/S_2^2$$

is then computed, and if $F \geq 2.89$, H_0 is rejected in favor of H_A . This plan should provide a test with power at least 0.80 when $\sigma_1^2/\sigma_2^2 = 5$. However, the sample sizes were obtained by interpolating in Table A6, and it is of interest to verify that the proposed plan is adequate.

This is accomplished by simulation analysis as follows.

1. Generate a random deviate x_1 from the chi-square distribution with $\nu_1 = 13$ degrees of freedom.
2. Generate a random deviate x_2 from the chi-square distribution with $\nu_2 = 6$ degrees of freedom.
3. Apply Equation 11.10 to compute the random deviate

$$x = \frac{x_1/13}{x_2/6} = \frac{6 x_1}{13 x_2}$$

Then compute the deviate

$$y = \frac{\sigma_1^2}{\sigma_2^2} x = 5x$$

4. Steps 1, 2, and 3 are repeated M times, where M is chosen to be large.

5. Count the number, say m , of deviates y that are greater than or equal to 2.89. The power of the test is then approximately $1 - \beta \cong m/M$.

These steps were carried out on a computer with $M = 5000$. There were $m = 4003$ deviates y which were greater than or equal to 2.89. Thus, the power of the test is approximately

$$1 - \beta \cong 4003/5000 = 0.8006$$

when $\sigma_1^2/\sigma_2^2 = 5$. This verifies that the proposed plan meets the power requirement.

11.3.7 The Multivariate Normal Distribution

Some stochastic modeling applications involve multiple random variables that may be correlated. In simulating such a model, it is necessary to

generate correlated random deviates. Section 2.6.3 discusses the general case of k jointly distributed random variables and provides definitions of the k -element mean vector $\underline{\mu}$ and the k -by- k symmetric variance-covariance matrix $\underline{\Sigma}$. One of the joint distributions used extensively in simulation modeling is the multivariate normal distribution. The general k -variate normal distribution is not explicitly defined in any of the previous chapters, but a detailed presentation of the two-variate (or bivariate) normal is given in Section 2.6.4.2.

To generate a k -element random vector \underline{x} from the k -variate normal distribution with mean vector $\underline{\mu}$ and variance-covariance matrix $\underline{\Sigma}$, the following steps are required:

1. Obtain a k -by- k matrix \underline{T} such that $\underline{T}'\underline{T} = \underline{\Sigma}$. The matrix \underline{T} can be obtained by performing a Cholesky square root factorization of the matrix $\underline{\Sigma}$. Graybill (1976) gives detailed instructions for performing a square root factorization.

2. Generate k independent random deviates, z_1 through z_k , from the standard normal distribution, and arrange them in a k -element vector \underline{z} . That is,

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_k \end{bmatrix}$$

3. Compute the k -element vector \underline{x} from the formula

$$\underline{x} = \underline{T}'\underline{z} + \underline{\mu} \quad (11.11)$$

The resulting vector \underline{x} is the equivalent of a random vector from the k -variate normal distribution with mean vector $\underline{\mu}$ and variance-covariance matrix $\underline{\Sigma}$. If steps 2 and 3 are repeated N times, a random sample of N k -variate vectors from this distribution is generated.

11.3.8 The Binomial Distribution

The binomial distribution is described in Section 2.4.2. A random deviate x from a binomial distribution with parameters n and p is generated by selecting n unit uniform random deviates and counting the number that are less than or equal to p . Symbolically, the steps of the procedure are as follows:

1. Generate n unit uniform random deviates r_i , $i = 1, 2, \dots, n$
2. Let $z_i = 1$ if $r_i \leq p$ and $z_i = 0$ if $r_i > p$
3. Then, $x = \sum_{i=1}^n z_i$

For example, let $n = 8$ and $p = 0.75$. Eight unit uniform random deviates were generated with values 0.12236, 0.95493, 0.21108, 0.04151, 0.58625, 0.58944, 0.98795, and 0.86085. The corresponding values of z_i are 1, 0, 1, 1, 1, 0, 0, and the resulting binomial random deviate is $x = 5$.

11.3.9 Functions of Random Variables

Sections 2.7 and 3.6 address the topic of functions of random variables. In general, interest is focused on a random variable Y which is a known mathematical function of one or more random variables. This is expressed as

$$Y = h(X_1, X_2, \dots, X_k) \quad (11.12)$$

where the individual X_i are random variables from perhaps different distributions, and h is a known mathematical function of the variables X_i .

Because the distributions of the individual random variables X_i are known, a random deviate from the (perhaps unknown) distribution of Y is generated as follows. First, generate one random deviate x_i from each of the individual distributions. Then substitute the random deviates x_i for the random variables X_i in Equation 11.12. The resulting quantity, given by

$$y = h(x_1, x_2, \dots, x_k) \quad (11.13)$$

is a random deviate from the distribution of Y . Repeating this procedure N times provides the equivalent of a random sample of size N from the distribution of Y .

Because the exact distribution of Y is usually unknown, the purpose of conducting a simulation study may be to characterize this distribution. A histogram, together with the sample mean and sample standard deviation, provides an initial characterization which may be adequate for some applications. Additional statistics (such as percentiles and higher sample moments) provide a more thorough characterization which may be required, for example, to "fine-tune" a hypothesis testing procedure. A characterization of the distribution of Y could indicate, for example, whether or not the normal distribution provides a satisfactory approximation for a given application. The goodness-of-fit tests presented in Chapter 9 can be applied to simulated data for this purpose.

Note that correlations among any of the random variables X_i must be taken into account when generating the deviates x_i . This is accomplished by using multivariate techniques such as the one presented in Section 11.3.7.

Example 11.8 In Example 2.11, the amount of ^{235}U in a container is measured by determining the net weight of the powder and drawing a sample of powder for analysis of percent uranium and percent ^{235}U . The following random variables are of interest:

X_1 = net weight of UO_2 powder with mean μ_1 and measurement variance σ_1^2

X_2 = ratio of uranium to UO_2 (i.e., proportion uranium) with mean μ_2 and measurement variance σ_2^2

X_3 = ratio of ^{235}U to uranium with mean μ_3 and measurement variance σ_3^2

$Y = X_1X_2X_3$ = amount of ^{235}U in the container with mean μ_Y and variance σ_Y^2

The following parameter values are assumed:

$$\begin{array}{ll} \mu_1 = 20.0 \text{ kg} & \sigma_1 = 0.05 \\ \mu_2 = 0.876 & \sigma_2 = 0.001 \\ \mu_3 = 0.0425 & \sigma_3 = 0.0002 \end{array}$$

It is further assumed that X_1 , X_2 , and X_3 are uncorrelated, so that all covariances are zero. Then, using the Taylor series approximation method, described in Section 2.7.2, the approximate mean and variance of Y are computed in Example 2.11 to be

$$\mu_Y \cong 0.7446 \text{ kg } ^{235}\text{U}$$

and

$$\sigma_Y \cong 0.0041 \text{ kg } ^{235}\text{U}$$

Suppose that in addition to the previous assumptions, it can be assumed that X_1 , X_2 , and X_3 follow normal distributions. It is of interest to investigate the distribution of Y and to obtain approximations of μ_Y and σ_Y^2 by simulation analysis. To do this, the following steps are necessary:

1. Generate three independent random deviates x_1 , x_2 , and x_3 , where x_1 is from the normal distribution with mean 20.0 and standard deviation 0.05; x_2 is from the normal distribution with mean 0.876 and standard deviation 0.001; and x_3 is from the normal distribution with mean 0.0425 and standard deviation 0.0002.

2. Compute the deviate

$$y = x_1x_2x_3$$

3. Steps 1 and 2 are repeated N times, where N is chosen to be quite large.

4. Use the resulting N deviates to compute a sample mean \bar{Y} and a sample standard deviation S .

5. Because the distribution of Y is of interest, a histogram can be plotted using the N deviates.

These steps were carried out on a computer with $N = 5000$. The resulting estimates of μ_Y and σ_Y are

$$\bar{Y} = 0.74455$$

and

$$S = 0.004079$$

These values agree with the approximations obtained in Example 2.11, where no particular distributions were assumed for X_1 , X_2 , and X_3 . A histogram constructed from the N generated y values follows.

| Middle of interval | Number of observations† |
|--------------------|-------------------------|
| 0.730 | 1 * |
| 0.732 | 12 * |
| 0.734 | 34 ** |
| 0.736 | 104 ***** |
| 0.738 | 258 ***** |
| 0.740 | 567 ***** |
| 0.742 | 832 ***** |
| 0.744 | 904 ***** |
| 0.746 | 888 ***** |
| 0.748 | 703 ***** |
| 0.750 | 416 ***** |
| 0.752 | 185 ***** |
| 0.754 | 70 **** |
| 0.756 | 24 ** |
| 0.758 | 2 * |

†Each * represents 20 observations, so that each row of *s represents the number of observations rounded up to the nearest multiple of 20.

Suppose that it is of interest to further characterize the distribution of Y by computing sample percentiles. The $100p^{\text{th}}$ percentile can be easily determined by arranging the N deviates in ascending order and then locat-

ing the N_p^{th} ordered deviate. To illustrate this, consider determining the 5th and 95th percentiles of the distribution of Y . When the 5000 deviates are arranged in ascending order, the 5th percentile is the $5000(0.05)^{\text{th}}$ or 250th ordered deviate, which is 0.73797. The 95th percentile is the $5000(0.95)^{\text{th}}$ or 4750th ordered deviate, which is 0.75117.

When the distribution of Y is unknown and there is some doubt about applicability of the normal theory techniques, simulation can be used, for example, to characterize the distributions of the sample mean and sample standard deviation under different sampling schemes. This would aid, for example, in determining required sample sizes and rejection limits for hypothesis tests about μ_Y and σ_Y .

REFERENCES

- Atkinson, A. C., and M. C. Pearce, 1976, The Computer Generation of Beta, Gamma, and Normal Random Variables, *J. Royal Stat. Soc.*, A139: 431-448.
- Bobachevsky, I. O., M. E. Johnson, and M. L. Stein, 1986, Generalized Simulated Annealing for Function Optimization, *Technometrics*, 28: 209-217.
- Box, G. E. P., and M. E. Muller, 1958, A Note on the Generation of Normal Deviates, *Anal. Math. Stat.*, 29: 610-613.
- Bratley, P., B. L. Fox, and L. E. Schrage, 1983, *A Guide to Simulation*, Springer-Verlag, New York.
- Emshoff, J. R., and R. L. Sisson, 1970, *Design and Use of Computer Simulation Models*, Macmillan Publishing Co., Inc., New York.
- Fishman, G. S., 1973, *Concepts and Methods in Discrete Event Digital Simulation*, John Wiley & Sons, New York.
- , and P. J. Kiviat, 1967, Digital Computer Simulation: Statistical Considerations, Memorandum RM-5387-PR, The Rand Corporation, Santa Monica, California.
- Graybill, F. A., 1976, *Theory and Application of the Linear Model*, Duxbury Press, Belmont, California.
- Hunter, J. S., and T. H. Naylor, 1970, Experimental Designs for Computer Simulation Experiments, *Manage. Sci.*, 16(7): 422-434.
- Kennedy, W. J., and J. E. Gentle, 1980, *Statistical Computing*, Marcel Dekker, Inc., New York.
- Kinderman, A. J., and J. G. Ramage, 1976, Computer Generation of Normal Random Variables, *J. Amer. Stat. Assoc.*, 71: 893-896.
- Kliejnen, J. P. C., 1975, *Statistical Techniques in Simulation*, Marcel Dekker, Inc., New York.
- Kolata, G., 1986, What Does it Mean to be Random?, *Science*, 231: 1068-1070.
- Law, A. M., and D. Kelton, 1982, *Simulation Modeling and Analysis*, McGraw-Hill Inc., New York.
- Marsaglia, G., 1968, Random Numbers Fall Mainly in the Planes, *Proc. Natl. Acad. Sci.*, 61: 25-28.
- , and T. A. Bray, 1964, A Convenient Method for Generating Normal Variables, *SIAM Review*, 6: 260-264.
- Meier, R. C., W. T. Newell, and H. L. Pazer, 1969, *Simulation in Business and Economics*, Prentice-Hall, New Jersey.

- Myers, R. H., 1976, *Response Surface Methodology*, Virginia Polytechnic Institute, Blacksburg, Virginia.
- Naylor, T. H., 1971, *Computer Simulation Experiments with Models of Economic Systems*, John Wiley & Sons, Inc., New York.
- , J. L. Balintfy, D. S. Burdick, and K. Chu, 1966, *Computer Simulation Techniques*, John Wiley & Sons, Inc., New York.
- Rubinstein, R. Y., 1981, *Simulation and the Monte Carlo Method*, John Wiley & Sons, Inc., New York.
- Schmidt, J. W., and R. E. Taylor, 1970, *Simulation and Analysis of Industrial Systems*, R. D. Irwin, Inc., Homewood, Illinois.
- The Rand Corporation, 1955, *A Million Random Digits with 100,000 Normal Deviates*, Free Press of Glencoe, Glencoe, Illinois.

CHAPTER 12

Elements of Decision Theory

12.0 INTRODUCTION

Decision theory combines elements of the theory of games and statistical hypothesis testing and is based on the principle that statistical procedures should be judged by their consequences. As developed by Wald (1950) and others, decision theory is a complex theoretical tool for mathematical statisticians. However, the basic principles of decision theory can be applied to the nuclear material accounting area without the necessity of invoking its full mathematical generality. This chapter is limited to an explanation, in a safeguards context, of how decision theory can be used to reach a decision in the face of uncertainty. Because of its limited scope, this chapter does not consider the relationship between decision theory and game theory or hypothesis testing. Introductory expositions of decision theory can be found in Chernoff and Moses (1959), Ferguson (1967), and Pratt, Raiffa, and Schlaifer (1964). Davis (1970) provides a nontechnical introduction to game theory. Luce and Raiffa (1958) present the central ideas and results of game theory and related decision-making models unencumbered by technical mathematical details. Blackwell and Girshick (1954) wrote a graduate level text that presents statistical decision theory from a game-theoretic point of view.

Decision theory has been applied to nuclear material management problems in a game theory context. Tingey (1979) observes that the commonly used approach to nuclear material inventory control tends to ignore the cost consequences of such occurrences as a false alarm or a process shutdown. He defines the total risk associated with an action as a function of measurement variability and the cost of an incorrect result. He then chooses a sampling plan that optimizes a combination of total risk and probability of detecting a true diversion. Opelka et al. (1979) review the conclusions of a peer review group convened to evaluate the applicability of game theory to the analysis of inventory differences. Moglewer (1978) uses game theory to validate the inventory difference concept for detecting unauthorized diversion of nuclear materials. He finds that game theory leads to strategies that are quite different from strategies based on classical statistical methods. Avenhaus (1978) discusses how a diverter of nuclear materials might choose a strategy for deciding which items to falsify in order to maximize the probability of not being detected. He then

discusses how the inspector should select samples in order to minimize the diverter's maximum probability of nondetection. Seaver et al. (1984) investigates the feasibility of applying game theory to the analysis of inventory differences.

This chapter begins with a short discussion of the distinction between decision theory and decision analysis, as used in the safeguards literature. This is followed by a section on definitions. Most of the chapter is a detailed discussion of an illustrative example that explains the principles of decision theory.

12.1 DECISION ANALYSIS

The term **decision analysis** has appeared in the nuclear material management literature since the late 1950s. More recent citations are Hakkila et al. (1977), Shipley (1978, 1980), and Bennett (1982a). Decision analysis is to be distinguished from the statistical decision theory introduced in this chapter. Decision theory (but not decision analysis) is discussed in many statistical textbooks.

Decision analysis as used in the nuclear material management literature is an organized plan for collecting data, analyzing it, and arriving at conclusions (Bennett, 1982b). It differs from decision theory in that the tools described in this chapter, such as loss functions and strategies, are not commonly used. The tools of decision analysis are hypothesis testing, data analysis, and sampling. This book provides the necessary background for decision analysis but does not discuss plans of action for evaluating a problem situation and determining a strategy for the resolution of that problem situation. In the applications chapters of this book (Chapters 13 through 17), many examples of such plans of action are discussed. Most of these are rather obvious combinations of the statistical principles presented in the previous chapters. The list of additional reading materials for this chapter includes papers that discuss this type of decision analysis.

12.2 DEFINITIONS

There are two kinds of uncertainty that need to be considered here. One kind is due to randomness or statistical phenomena. This is the randomness that explains how dice fall, the way a flipped coin falls, or which hands are dealt in a game of cards. This randomness is described with statistical distributions, such as those discussed in Chapter 2. Most of this text is devoted to the statistical aspects of this kind of uncertainty. For example, some measurement errors, production lot variability, and sampling errors are of this kind. The other kind of uncertainty arises when it is not known which statistical laws of randomness actually apply to a given situation. In Chapter 6, regression model validation and other approaches to finding the correct model are discussed as methods to reduce the magni-

tude of this second kind of uncertainty. In statistical terminology, this second kind of uncertainty is "ignorance of the state of nature." In nuclear material safeguards, the state of nature is whether or not a diversion has occurred and, if so, what diversion strategy was used. Hypothesis testing aids in deciding whether or not a hypothesized state of nature is correct in the face of the uncertainty caused by randomness in the available information (Section 16.2.2).

In hypothesis testing there are two **available actions**, either to reject or not to reject the hypothesis. In the decision theory approach described in this chapter, there may be many available actions or decisions that can be made. The consequences of taking an action depend upon the true state of nature. The **loss function** measures the cost or gain of taking specified actions for all possible states of nature. The loss function is typically displayed in a **loss table**, where all possible actions are listed on one axis and all possible states of nature on the other. The entries in the table are the cost of each combination of action and state of nature. Given a loss table, it would be easy to select the best action *if* the state of nature was known. To gain information to assess the state of nature, an **experiment** may be performed which yields data. A **strategy** tells how to use such data to make a **decision**.

12.3 AN ILLUSTRATIVE EXAMPLE

The principles of decision theory are illustrated by detailed discussion of the following example.

12.3.1 Statement of the Problem

One part of the nuclear material inventory of any fuel fabrication facility is the scrap and contaminated waste generated during the manufacture of fuel rods. These scrap and waste materials are usually stored in drums, which are used to transport these materials for disposal. A facility may want to dispose of these drums in the most economical way. Suppose that drums with a surface exposure rate of 5 R/h or less can be sent to a low-level disposal site at a cost of \$80 per drum. Drums with a surface exposure rate over 5 R/h must go to a high-level waste site at a cost of \$350 per drum. Each drum must be measured, and a decision must be made as to where to send it. If a high-level drum is sent to a low-level site, there is a 50% chance that the error will be detected and a fine of \$450 will be incurred. Suppose further that a simple radiation meter can be used for scanning the surface of the drums, or the drums can be sent to a contractor for accurate scanning at a cost of \$50 per drum. A strategy must be devised which minimizes disposal costs.

This is not yet a complete statement of the problem, but the situation described so far, in terms of its decision theory structure, entails two states

of nature: each drum either contains high-level waste or it contains low-level waste. There are three available actions: a drum can be sent to a low-level waste site, it can be sent for accurate measurement, or it can be sent to a high-level waste site. Also, some information is available on costs associated with the various actions. Thus, a table of the loss function for this situation can be constructed. Denote by N_1 the state of nature of a drum containing low-level waste and by N_2 the state of nature of a drum containing high-level waste. Denote by A_1 the action of sending a drum to a low-level site, by A_2 the action of sending a drum for accurate measurement, and by A_3 the action of sending a drum to a high-level site. Table 12.1 describes the loss function.

TABLE 12.1
Loss Function for Waste
Disposal (Dollars)

| State of nature | Action | | |
|--------------------|--------|-------|-------|
| | A_1 | A_2 | A_3 |
| N_1 (low-level) | 80 | 130 | 350 |
| N_2 (high-level) | 440 | 400 | 350 |

The entries in this table are determined as follows. If a drum is truly low-level and it is sent to a low-level site, the cost is \$80 as described above. If a low-level drum is sent for accurate analysis, the cost is \$50 for the analysis plus \$80 for disposal, a total of \$130. A low-level drum sent to a high-level site will cost the same as a high-level drum, which is \$350. If a high-level drum is accurately analyzed first, a \$50 analysis cost is added for a total of \$400. If a high-level drum is sent to a low-level site, an adjustment must be made for the 50% chance of getting caught, the fine, and the cost of correct disposal. To estimate this loss, the principle of expectation described in Chapter 2 is applied. The high-level drum will be accepted for \$80 at the low-level site with probability 0.50. The probability is also 0.50 of having to pay the \$450 fine plus \$350 for correct disposal. The expected loss is then $0.5 (\$80) + 0.5 (\$450 + \$350) = \440 .

Suppose that the choice of actions is based upon the reading from the simple radiation meter applied to the surface of each drum. These readings, of course, contain random measurement error. For now, assume there is no prior knowledge of drum contents; such knowledge is used in more sophisticated analyses. Assume that a drum will be sent to a low-level site if the radiation meter registers less than 4 R/h and to a high-level site if the meter registers above 6 R/h. If the reading is between 4 and 6 R/h, the drum will be sent for accurate analysis. The probability of any specific

reading on the meter depends upon the true state of nature. By having samples of all kinds of drums accurately analyzed, estimates of the probabilities of the various readings can be obtained. Suppose that the results of such an exercise are given in Table 12.2. Here, X_1 represents a reading less than 4 R/h, X_2 a reading between 4 and 6 R/h, and X_3 a reading above 6 R/h. The probabilities in this table could have also been estimated from historical records, simulations of the waste production, or even educated guesses of drum contents.

TABLE 12.2

Probabilities of Observing a Radiation Meter Reading of X_j When N_i Is the True State of Nature

| State of nature | Reading | | |
|-----------------|---------|-------|-------|
| | X_1 | X_2 | X_3 |
| N_1 | 0.65 | 0.25 | 0.10 |
| N_2 | 0.10 | 0.30 | 0.60 |

12.3.2 Choosing a Strategy

The next step is to decide how to respond to the radiation meter reading. This response plan is called the **strategy**. There are three possible actions: A_1 , A_2 , and A_3 . There are also three possible readings: X_1 , X_2 , and X_3 . A strategy is a rule that determines which action, A_i , to take based on the observed reading, X_j . Many strategies are obviously ridiculous or counterproductive, but for completeness all must be listed. In this example, there are 27 possible strategies; they are denoted by S_1 through S_{27} and are displayed in Table 12.3.

Strategy 1 ignores the data and sends everything to the low-level waste site. Strategy 27 also ignores the data and sends everything to the high-level site. Strategy 6 intuitively appears to be the best, but may actually cost more than some other strategy because of the cost of accurate analysis. Strategy 9 avoids the cost of an accurate analysis by sending questionable drums directly to the high-level waste site.

The next step is to compute the expected loss associated with each strategy. Information from the preceding three tables is used with the principles of expectation discussed in Chapter 2. The expected loss for a strategy is defined as the **risk** associated with using that strategy. As an illustration, consider Strategy 5. For this strategy, action A_1 is taken if X_1 is observed, and action A_2 if either X_2 or X_3 is observed. When N_1 is the state of nature, X_1 is observed with probability 0.65 and X_2 or X_3 with probability $0.25 + 0.10 = 0.35$. With this strategy, action A_3 is never

TABLE 12.3
Possible Strategies

| Possible observations | Strategies | | | | | | | | |
|-----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | S ₁ | S ₂ | S ₃ | S ₄ | S ₅ | S ₆ | S ₇ | S ₈ | S ₉ |
| X ₁ | A ₁ | A ₁ | A ₁ | A ₁ | A ₁ | A ₁ | A ₁ | A ₁ | A ₁ |
| X ₂ | A ₁ | A ₁ | A ₁ | A ₂ | A ₂ | A ₂ | A ₃ | A ₃ | A ₃ |
| X ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ |
| | S ₁₀ | S ₁₁ | S ₁₂ | S ₁₃ | S ₁₄ | S ₁₅ | S ₁₆ | S ₁₇ | S ₁₈ |
| X ₁ | A ₂ | A ₂ | A ₂ | A ₂ | A ₂ | A ₂ | A ₂ | A ₂ | A ₂ |
| X ₂ | A ₁ | A ₁ | A ₁ | A ₂ | A ₂ | A ₂ | A ₃ | A ₃ | A ₃ |
| X ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ |
| | S ₁₉ | S ₂₀ | S ₂₁ | S ₂₂ | S ₂₃ | S ₂₄ | S ₂₅ | S ₂₆ | S ₂₇ |
| X ₁ | A ₃ | A ₃ | A ₃ | A ₃ | A ₃ | A ₃ | A ₃ | A ₃ | A ₃ |
| X ₂ | A ₁ | A ₁ | A ₁ | A ₂ | A ₂ | A ₂ | A ₃ | A ₃ | A ₃ |
| X ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ | A ₁ | A ₂ | A ₃ |

taken. Thus, for state of nature N_1 , actions A_1 , A_2 , and A_3 (which give losses of 80, 130, and 350, respectively) are taken with probabilities 0.65, 0.35, and 0.0, respectively. The expected loss is then $0.65(80) + 0.35(130) + 0.0(350) = 97.5$. The expected losses for each strategy and each state of nature are calculated in this manner and are displayed in Table 12.4.

TABLE 12.4
Expected Loss for Each Strategy and State of Nature

| State of nature | Strategy | | | | | | | | |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | S ₁ | S ₂ | S ₃ | S ₄ | S ₅ | S ₆ | S ₇ | S ₈ | S ₉ |
| N ₁ | 80.0 | 85.0 | 107.0 | 92.5 | 97.5 | 119.5 | 147.5 | 152.5 | 174.5 |
| N ₂ | 440.0 | 416.0 | 386.0 | 428.0 | 404.0 | 374.0 | 413.0 | 389.0 | 359.0 |
| | S ₁₀ | S ₁₁ | S ₁₂ | S ₁₃ | S ₁₄ | S ₁₅ | S ₁₆ | S ₁₇ | S ₁₈ |
| N ₁ | 112.5 | 117.5 | 139.5 | 125.0 | 130.0 | 152.0 | 180.0 | 185.0 | 207.0 |
| N ₂ | 436.0 | 412.0 | 382.0 | 424.0 | 400.0 | 370.0 | 409.0 | 385.0 | 355.0 |
| | S ₁₉ | S ₂₀ | S ₂₁ | S ₂₂ | S ₂₃ | S ₂₄ | S ₂₅ | S ₂₆ | S ₂₇ |
| N ₁ | 255.5 | 260.5 | 282.5 | 268.0 | 273.0 | 295.0 | 323.0 | 328.0 | 350.0 |
| N ₂ | 431.0 | 407.0 | 377.0 | 419.0 | 395.0 | 365.0 | 404.0 | 380.0 | 350.0 |

The algorithm for these computations can be visualized as follows. Make two copies of Table 12.3; in the first copy replace the A_i 's with their values for the first state of nature found in Table 12.1, and in the second copy use the values for the second state of nature in Table 12.1. Of course, in the general problem there would be n states of nature, and n copies of Table 12.3 would be made. Next, replace the X_j 's in these two tables by their values found in Table 12.2, again using the appropriate state of nature. Then multiply each entry within the table by the X_j value for its row, and sum down the column for each strategy. These sums are the numbers found in Table 12.4.

Table 12.4 gives the expected loss for every possible combination of strategy and state of nature. Accordingly, the choice of strategy should depend only on the numbers in Table 12.4. But how is this choice to be made? It is clear from Table 12.4 that the best strategy depends on the state of nature. If N_1 is true, S_1 minimizes the expected loss and if N_2 is true, S_{27} minimizes the expected loss. However, if S_1 is used and N_2 is true, or if S_{27} is used and N_1 is true, the expected loss will be maximized. In other words, the best strategy for one state is the worst strategy for the other state.

A number of criteria have been suggested for choosing a strategy from Table 12.4. The **minimax criterion** chooses the strategy which minimizes the maximum expected loss. The criterion assumes that the worst will always happen; that is, whatever strategy is chosen, the true state of nature will be that state which maximizes the risk. Hence, each strategy is evaluated by the maximum expected loss over the states of nature, and the minimax criterion seeks to minimize this loss. Referring to Table 12.4, if strategy S_1 is chosen, the maximum expected loss is 440.0, if strategy S_2 is chosen, the maximum expected loss is 416.0, and so on. The maximum expected loss is minimized (at 350.0) if S_{27} is chosen. Hence, S_{27} is the **minimax strategy** and the **minimax risk** is 350.0. From Table 12.3, S_{27} corresponds to always taking action A_3 . Always sending a drum to a high-level site therefore minimizes the maximum expected loss regardless of the state of nature.† In general, the minimax strategy consists of always taking a particular one of the available actions; that is, no other action is ever taken.

Because the minimax criterion is a conservative criterion that assumes the worst will always happen, it ignores the information available in the radiation meter readings. It also ignores any prior information about the relative likelihood of the states of nature. By taking this likelihood into account, it is possible to reduce the average risk to a value below the minimax risk.

†This conclusion follows directly from Table 12.1 without the necessity of constructing Table 12.4.

Suppose that state N_1 and state N_2 are judged to be equally likely. Then the risk of using any strategy is simply the average of the expected losses for N_1 and N_2 from Table 12.4. These average losses are displayed in Table 12.5. From Table 12.5, the best strategy (i.e., the strategy that minimizes the risk) is S_3 with a risk of 246.50. This strategy is to send low-level and also questionable drums to the low-level site and avoid the cost of accurate analysis by taking a chance on being caught; high-level drums are sent to the high-level site. Strategy S_{27} , sending everything to the high-level site, is expected to cost 1.42 times as much as strategy S_3 , so the conservative minimax strategy is expected to be 42% more expensive.

TABLE 12.5
Risk for Each Strategy if N_1 and N_2 Are Equally Likely

| Strategy | Risk | Strategy | Risk | Strategy | Risk |
|----------|--------|----------|--------|----------|--------|
| 1 | 260.00 | 10 | 274.25 | 19 | 343.25 |
| 2 | 250.50 | 11 | 264.75 | 20 | 333.75 |
| 3 | 246.50 | 12 | 260.75 | 21 | 329.75 |
| 4 | 260.25 | 13 | 274.50 | 22 | 343.50 |
| 5 | 250.75 | 14 | 265.00 | 23 | 334.00 |
| 6 | 246.75 | 15 | 261.00 | 24 | 330.00 |
| 7 | 280.25 | 16 | 294.50 | 25 | 363.50 |
| 8 | 270.75 | 17 | 285.00 | 26 | 354.00 |
| 9 | 266.75 | 18 | 281.00 | 27 | 350.00 |

What about the “logical” strategy, S_6 , of sending the questionable drums for accurate analysis? The risk of S_6 is 246.75, just 0.1% higher than the risk of the best strategy; this raises the question of whether this slight extra loss should be incurred in order to avoid the repercussions resulting from having drums rejected by the low-level dump site. Such additional considerations could be included in the analysis by increasing the cost associated with having a drum rejected.

12.3.3 Weighting of Losses and Bayes Strategies

Averaging the losses over the states of nature is not the only way to derive a single value of risk for each strategy. Note that, with two states of nature, averaging is equivalent to a weighted sum in which each state of nature has a weight of 0.5. Any combination of weights that sum to unity would result in a weighted average over the states of nature. For example, if it is judged that the fuel fabrication plant in the example generates twice as many low-level drums as high-level drums, and if the process is

stable, the weights should be $2/3$ for state N_1 and $1/3$ for state N_2 . Using these weights to sum over the columns of Table 12.4 yields the risks given in Table 12.6.

TABLE 12.6
Risk for Each Strategy if
 N_1 Weight = $2/3$ and N_2 Weight = $1/3$

| Strategy | Risk | Strategy | Risk | Strategy | Risk |
|----------|-------|----------|-------|----------|-------|
| 1 | 200.0 | 10 | 220.3 | 19 | 314.0 |
| 2 | 195.3 | 11 | 215.7 | 20 | 309.3 |
| 3 | 200.0 | 12 | 220.3 | 21 | 314.0 |
| 4 | 204.3 | 13 | 224.7 | 22 | 318.0 |
| 5 | 199.7 | 14 | 220.0 | 23 | 313.7 |
| 6 | 204.3 | 15 | 224.7 | 24 | 318.3 |
| 7 | 236.0 | 16 | 256.3 | 25 | 350.0 |
| 8 | 231.0 | 17 | 251.7 | 26 | 345.3 |
| 9 | 236.0 | 18 | 256.3 | 27 | 350.0 |

With these weights, the best strategy is S_2 , to send drums with low and questionable readings to the low-level disposal site and to send all drums with high readings for accurate analysis. The "logical" strategy, S_6 , now has an average loss 5% higher than the minimum loss achievable with S_2 .

In general, any prior judgement about the relative likelihoods of N_1 and N_2 corresponds to nonnegative weights on N_1 and N_2 which sum to unity. This set of weights forms a **prior distribution** on the states of nature. A strategy (there might possibly be more than one) whose weighted average loss is a minimum is called a **Bayes strategy** with respect to the given prior distribution, and the minimum weighted average loss is called the **Bayes risk** for the given prior distribution.

If the prior judgement about the states of nature is accurate, then the optimal strategy is the Bayes strategy for the given prior distribution. However, if the prior judgement is inaccurate, then the corresponding Bayes strategy may not be optimal, i.e., its risk may be larger than the risk of using a Bayes strategy corresponding to the correct prior distribution. For example, suppose it is incorrectly believed that N_1 and N_2 are equally likely and, as a consequence, S_3 is chosen. If N_1 is really twice as likely to occur as N_2 , then Table 12.6 applies and S_2 is the corresponding Bayes strategy. If S_3 is used, its average loss of 200.0 is 2.4% higher than the minimum Bayes risk achievable with S_2 . Similarly, if the minimax assumption holds and N_2 always occurs, then the expected loss due to using S_3 is 386.0, or 10.3% higher than the minimax risk if the optimal strategy S_{27} had been used. Hence, before adopting a Bayes strategy based

on a particular prior distribution, the validity of the prior distribution should be evaluated and a sensitivity analysis of the consequences of using an inaccurate prior distribution should be performed.

A typical source of prior information is the records that are kept about drums; however, there is often no assurance that the records are correct. If information on individual drums is available, the methodology presented in this chapter could be applied to each drum rather than to all drums as a group. This would, of course, require a substantial increase in the computational effort.

It should be noted that simple averaging (equal weights) assumes that all states of nature are equally probable. This weighting scheme has a special name; it is called the **noninformative prior distribution**, and it is often used when no information is available about the states of nature.

12.3.4 Admissible Strategies

For all the prior distributions discussed so far, the best or close-to-best strategies are those indexed by the numbers 1 through 6. In some sense, these strategies seem to be better than all others, regardless of the prior distribution. This is investigated further below.

Because for the above example there are two expected losses for each strategy, one corresponding to each state of nature, a strategy can be represented by a point on a graph with coordinates representing the expected losses. Let the abscissa be the expected loss when N_1 is the state of nature and let the ordinate be the expected loss when N_2 is the state of nature. Plotting the data in Table 12.4 produces Figure 12.1.

A desirable strategy is one where both expected losses are small. As illustrated by Figure 12.1, no single strategy simultaneously has the smallest loss for both states of nature. It can also be seen from this figure that the strategies plotted in the upper right are never as good as others. For example, S_{14} should never be used because S_3 has a smaller expected loss regardless of which state of nature is true; we say that S_3 **dominates** S_{14} . Of the 27 strategies, the only ones that should be considered are $S_1, S_2, S_3, S_5, S_6, S_9, S_{15}, S_{18},$ and S_{27} . These strategies are called **admissible** because they are not dominated by any other strategy. The other strategies are called **inadmissible** because each is dominated by one or more admissible strategy.

So far, the number of strategies to be considered has been reduced from 27 to the 9 admissible strategies. The choice among the admissible strategies still depends upon the prior distribution, but only 9 weighted averages, rather than 27, need to be computed and compared.

12.3.5 Mixed Strategies

It is possible to reduce the number of strategies to be considered even further, to a subset of the set of admissible strategies. To see this, consider

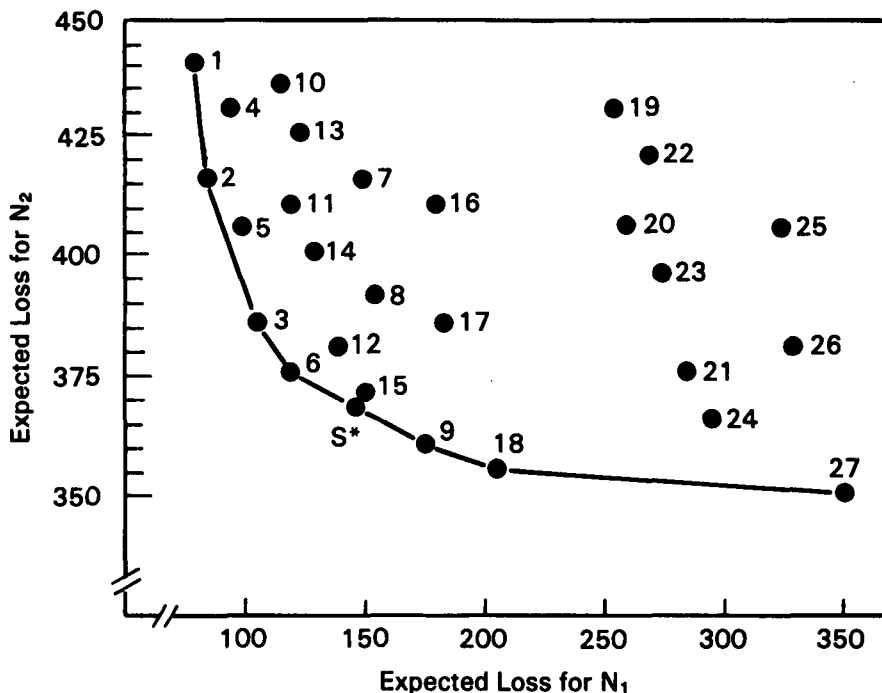


Figure 12.1 Plot of expected loss for N_1 against expected loss for N_2 for each of the 27 strategies and for the mixed strategy S^* (see Section 12.3.5).

the segmented curve connecting the seven “outside” points along the left-hand side and bottom of Figure 12.1. These correspond to the admissible strategies S_1 , S_2 , S_3 , S_6 , S_9 , S_{18} , and S_{27} . The points corresponding to the other two admissible strategies, S_5 and S_{15} , lie slightly inside this segmented curve. Even though S_5 and S_{15} are not dominated by any other strategies, intuitively it seems it is possible to do better. For example, consider the two strategies bracketing S_{15} , i.e., S_6 and S_9 . Define a new kind of strategy in which a coin is tossed. If it comes up heads, S_6 is chosen; if it comes up tails, S_9 is chosen.† This new kind of strategy is called a **mixed** or **randomized** strategy because it is a random mixture of two **pure** strategies. (A pure strategy is one that takes a unique action for each possible observation; the pure strategies are listed in Table 12.3.) For each state of nature, the expected loss is simply a weighted sum of the losses for S_6 and S_9 with weights $1/2$ and $1/2$. Thus, if N_1 is the state of nature, the new strategy has expected loss $(1/2)(119.5) + (1/2)(174.5) = 147.0$.

†From Table 12.3, this strategy is equivalent to the following: If X_1 is observed, choose A_1 ; if X_2 is observed, choose A_2 if the coin comes up heads or A_3 if the coin comes up tails; if X_3 is observed, choose A_3 .

Similarly, if N_2 is the state of nature, the expected loss is $(1/2)(374.0) + (1/2)(359.0) = 366.5$. If this new strategy, denoted by S^* , is plotted in Figure 12.1, it dominates S_{15} . Similarly, a weighted sum of S_2 and S_3 can be found that dominates S_5 . Thus, S_5 and S_{15} can be dropped from the set of admissible strategies, leaving only seven rather than nine strategies for consideration.

Mixed strategies have the somewhat disconcerting property that the action taken may depend upon a completely irrelevant random choice, e.g., a random number generator. However, the strategies must be judged only by their consequences, which are the risks or expected losses under both states of nature. With this criterion, mixed strategies can result in lower risks than some pure strategies.†

Note that mixed strategies fall on a straight line connecting the two pure strategies being mixed. The location along this line where the new strategy falls depends upon the mixing weights. If both weights are $1/2$, the new strategy is located in the middle of the line. If the weights are $2/3$ for S_6 and $1/3$ for S_9 , the new strategy is located $2/3$ of the way along the line and closer to S_6 . If all possible weights summing to unity are used, the resulting set of mixed strategies corresponds to the line segment connecting S_6 and S_9 . If line segments are drawn connecting all consecutive pairs of pure admissible strategies as in Figure 12.1, the resulting segmented curve represents all the admissible strategies.

12.3.6 Choosing a Bayes Strategy

For a given prior distribution of the states of nature, the problem is to choose a Bayes strategy from among the admissible strategies. Because every point on the segmented curve in Figure 12.1 corresponds to an admissible strategy, there are an infinite number of such strategies. Consequently, a graphical approach will be used.

The first step is to represent the overall weighted average loss for an arbitrary strategy S by a line. Let $L_1(S)$ represent the expected loss for strategy S and state of nature N_1 , and let $L_2(S)$ represent the expected loss for the same strategy and state of nature N_2 . These expected losses are displayed in Table 12.4 for the 27 pure strategies. Also, let $1 - w$ and w be the weights assigned to N_1 and N_2 , respectively, by the prior distribution of the states of nature. The weighted average loss or risk for Strategy S is then

$$R(S) = (1 - w)L_1(S) + wL_2(S) \quad (12.1)$$

The objective is to find the strategy which minimizes this weighted sum.

†However, it will be shown later in this section that Bayes strategies can always be chosen to be pure strategies.

Equation 12.1 can be rearranged in the form

$$L_2(S) = \frac{1}{w} R(S) - \frac{(1-w)}{w} L_1(S) \quad (12.2)$$

In this form, it can be seen that Equation 12.1 represents a family of straight lines, all with the same negative slope. These lines are parallel and differ only in their intercepts. Lines A and B on Figure 12.2 are two such lines for the family with $w = 1/2$. The segmented curve in Figure 12.2 labeled "Admissible Strategies" has been copied from Figure 12.1.

A Bayes strategy can be graphically determined by drawing a line on a loss plot (such as Figure 12.1 or 12.2) with slope $-(1-w)/w$ and then moving that line, while retaining its slope, up and to the right until it first touches the segmented curve of admissible strategies. The point (or points) first touched by the line is the set of Bayes strategies for the given value of w . For example, suppose that $w = 1/2$. In Figure 12.2, line A is drawn with slope $-(1 - 1/2)/(1/2) = -1.0$. Note that in Figure 12.2, the axes are broken and are scaled so that one unit on the vertical (N_2) axis is

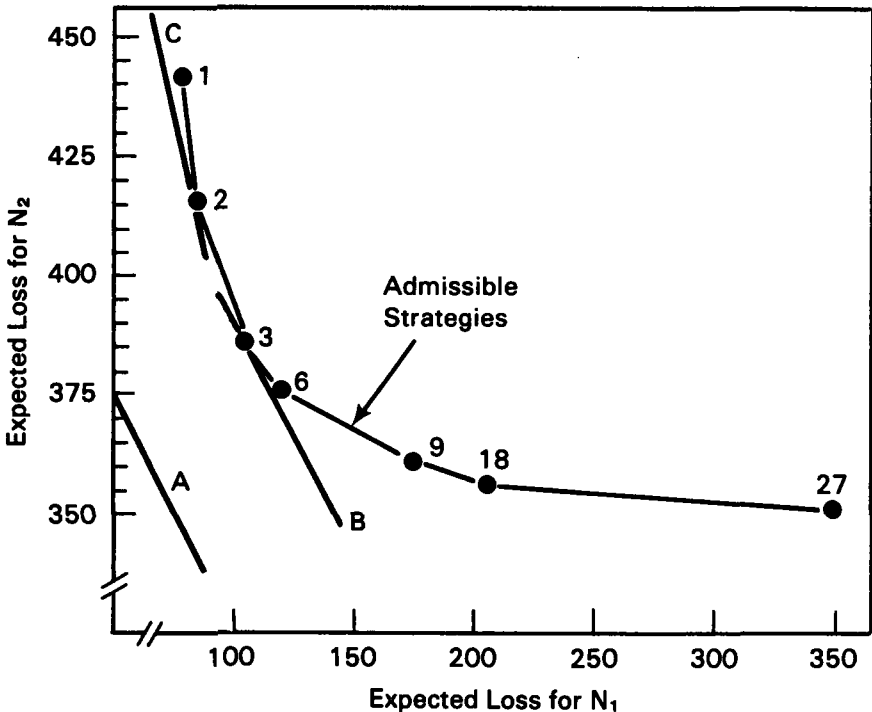


Figure 12.2 Loss lines and admissible strategies for states of nature N_1 and N_2 .

twice as long as one unit on the horizontal (N_1) axis. Now, move line A up and to the right, retaining its slope, until it first touches the curve of admissible strategies. The result is line B, which first touches the curve of admissible strategies at the point representing S_3 . Note that line B does not touch S_6 , although it is close. Hence, for $w = 1/2$, the unique Bayes strategy is S_3 . In the discussion of Table 12.4 it was noted that for $w = 1/2$, the best strategy is S_3 with an expected loss of 246.5, and that S_6 has an expected loss of 246.8, just 0.1% higher.

If w were 0.51 rather than 0.50, then line B in Figure 12.2 would pass through both S_3 and S_6 , and any mixed strategy on the line connecting S_3 and S_6 would be a Bayes strategy for $w = 0.51$. However, the pure strategies S_3 and S_6 would also be Bayes strategies. Hence, whether unique or not, a Bayes strategy can always be chosen to be a pure strategy.

Table 12.6 displays the results for the prior distribution leading to weights of 2/3 for N_1 and 1/3 for N_2 , and the best strategy is S_2 . Here, the value of w is 1/3, so the slope is $-(1 - w)/w = -(2/3)/(1/3) = -2.0$. Line C in Figure 12.2 is the Bayes strategy line with a slope of -2.0 . It shows graphically that S_2 is the preferred strategy under this prior distribution.

While the graphical display of losses and strategies is very helpful in understanding the relationship of strategies and is almost essential in picking the set of admissible strategies, it is generally easier to determine the Bayes strategy by actually computing the risks for the admissible strategies. This avoids the inaccuracies of graphical analysis. Plots like Figure 12.2 help to illustrate the effects of different prior distributions. A comparison of line B with $w = 1/2$ and line C with $w = 1/3$ shows that as drums of low-level waste become relatively more abundant, the Bayes strategy line becomes more vertical. Similarly, if high-level waste becomes relatively more abundant, the line becomes more horizontal. For example, the line on Figure 12.2 connecting points S_6 and S_9 has a slope of -0.273 , which corresponds to a 0.79 prior probability for the abundance of high-level drums. Also, as previously noted, the line connecting S_3 and S_6 corresponds to a 0.51 prior probability for high-level drums. Thus, we can conclude that S_6 , previously designated as the obvious or logical strategy, is the best strategy only if the relative abundance of high-level waste drums is between 51 and 79%.

From Equation 12.2, the intercept of the Bayes strategy line is equal to $R(S)/w$. If a Bayes strategy for a given w is denoted by S_w , then $R(S_w) = R_w$ is the Bayes risk and R_w is w times the intercept of the Bayes strategy line when it first touches the curve of admissible strategies. If the Bayes risk R_w is plotted against w for $0 \leq w \leq 1$, there will be a unique value of w , say w^* , which maximizes R_w . The weights $1 - w^*$ and w^* form the **least favorable distribution** on N_1 and N_2 because they maximize the Bayes risk. For the example, $w^* = 1$ and the least favorable

distribution places all its weight on N_2 . The corresponding Bayes strategy is S_{27} , which is the minimax strategy.

This result is true in general. The minimax strategy is the Bayes strategy for the least favorable distribution, i.e., the minimax risk is the maximum Bayes risk. Consequently, if the prior distribution for the states of nature is unknown, then the minimax strategy should be considered. The minimax strategy is appealing because it does not require any information about the prior distribution of the states of nature; thus, it is sometimes advocated as a reasonable alternative to the noninformative prior distribution.

12.3.7 The Regret Approach

An alternative to the loss function approach presented in Section 12.3.1 is the regret approach. The regret approach focuses on the incremental cost of not knowing the true state of nature instead of the absolute cost. For example, suppose that N_1 is true. Then from Table 12.1, the loss could be held to \$80 by taking action A_1 . If action A_2 is taken instead, then the incremental cost of not knowing that N_1 is true is $\$130 - \$80 = \$50$. The **regret function** is defined as the difference between the cost of the action taken and the cost of the optimal action for the given state of nature. The regret function for the example, based on Table 12.1, is given by Table 12.7.

TABLE 12.7
Regret Function for Waste Disposal (Dollars)

| State of nature | Action | | |
|--------------------|--------|-------|-------|
| | A_1 | A_2 | A_3 |
| N_1 (low-level) | 0 | 50 | 270 |
| N_2 (high-level) | 90 | 50 | 0 |

The **minimax regret strategy** is the strategy that minimizes the maximum regret. From Table 12.7, the minimax regret strategy is A_2 . This is different from the minimax strategy A_3 , which is based on the absolute losses given by Table 12.1. It can be argued that A_2 is preferable to A_3 because using A_2 involves an incremental cost of only \$50 over the cost of using the optimal strategy, regardless of which state of nature is true, while using A_3 leads to an incremental cost of \$270 if N_1 happens to be true.

Based on Table 12.7, the expected regrets for each strategy and state of nature can be tabulated exactly as in Table 12.4, and figures similar to Figures 12.1 and 12.2 can be drawn. In fact, the use of the regret function instead of the loss function leads to a simple shift of the points in Figure

12.1, so the segmented curve corresponding to the admissible strategies touches both axes. It then follows that the Bayes strategy which minimizes the expected regret for a given prior distribution is precisely the same as the Bayes strategy which minimizes the expected loss.

In contrast to the loss function approach, where the minimax strategy (S_{27}) is the same whether or not the radiation meter readings are made, the minimax regret strategy depends on the radiation meter readings. It can be shown that the minimax regret strategy is a mixture of S_3 and S_6 , with weights 0.63 and 0.37, respectively, and the minimax expected regret is 31.6. This is a 37% reduction from the minimax regret of 50 from Table 12.7. However, as for the loss function approach, the minimax regret is the same regardless of which state of nature is true. This is a general property of minimax strategies.

12.4 DISCUSSION

The example presented in Section 12.3 conveys the basic elements of the decision theory approach. There are six basic components of a typical statistical decision-making problem:

1. The state of nature.
2. The permissible actions.
3. The losses or consequences of the actions.
4. The experiment that yields data.
5. The strategies of how to respond to the data.
6. The prior probabilities of the states of nature.

It should be noted that some problems may have no experimental data, or the prior probabilities may not be known.

The tabular approach to examining all possibilities can be prohibitively complex in situations that involve many states of nature, many kinds of data observations, or many possible actions. The decision theory approach can be formulated and a solution obtained in a purely mathematical way. However, the tabular approach presented here has been found to be adequate for a wide variety of real problems.

This approach to decisions, in its rigorous mathematical form, can be extended to provide a mathematical foundation for all statistical hypothesis testing and parameter estimation methods. There is an obvious analogy between the null and alternate hypotheses and the two states of nature. The losses are analogous to specifying costs of making Type I and Type II errors (Section 4.1.2), and the actions and strategies are analogous to the test statistic and the critical values used in hypothesis testing.

REFERENCES

- Avenhaus, R., 1978, *Material Accountability—Theory, Verification, Applications*, John Wiley & Sons, Inc., New York.

- Bennett, C. A., 1982a, Statistics and Safeguards: From Control to Evaluation to Decision, *Proc. 1981 DOE Statistical Symposium*, BNL-51535, Brookhaven National Laboratory, Upton, New York.
- , 1982b, Introduction to Decision Analysis, in *Mathematical and Statistical Methods in Nuclear Safeguards*, Argentesi, F., R. Avenhaus, M. Franklin, and J. P. Shipley (Eds.), Harwood Academic Publishers, Chur, Switzerland.
- Blackwell, D., and M. A. Girshick, 1954, *Theory of Games and Statistical Decisions*, John Wiley & Sons, Inc., New York.
- Chernoff, H., and L. E. Moses, 1959, *Elementary Decision Theory*, John Wiley & Sons, New York.
- Davis, M. D., 1970, *Game Theory*, Basic Books, New York.
- Ferguson, T. A., 1967, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.
- Hakkila, E. A., J. W. Barnes, T. R. Canada, D. D. Cobb, S. T. Hsue, D. G. Langner, J. L. Parker, J. P. Shipley, and D. B. Smith, 1977, *Coordinated Safeguards for Materials Management in a Fuel Reprocessing Plant*, Vol. II Appendix, LA-6881, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.
- Luce, R. D., and H. Raiffa, 1958, *Games and Decisions*, John Wiley & Sons, Inc., New York.
- Moglewer, S., 1978, The Application of Game Theory for the Interpretation of Inventory Differences, *Nucl. Mater. Manage.* (Proc. Issue), 7: 298-304.
- Opelka, J. H., C. A. Bennett, A. J. Goldman, W. A. Higinbotham, J. L. Jaech, W. F. Lucas, and R. F. Lumb, 1979, Peer Review of Strategic Analysis of Material Accounting, *Nucl. Mater. Manage.* (Proc. Issue), 8: 507-514.
- Pratt, J. W., H. Raiffa, and R. Schlaifer, 1964, The Foundations of Decision Under Uncertainty: An Elementary Exposition, *J. Am. Stat. Assoc.*, 59: 353-375.
- Seaver, D. A., A. J. Goldman, W. H. Immerman, F. L. Crane, and M. S. Cohen, 1984, *Strategic Analysis for Safeguards Systems: A Feasibility Study*, NUREG/CR-3926, Vol. 1 and 2, U. S. Nuclear Regulatory Commission, Washington, D. C.
- Shipley, J. P., 1978, Decision Analysis for Nuclear Safeguards, *Proc. Symposium on Nuclear Safeguards Analysis*, American Chemical Society, Washington, D. C., 34-64.
- , 1980, Some Safeguards Approaches and Evaluation Methods, *Nucl. Mater. Manage.*, 7(4): 58-65.
- Tingey, F. H., 1979, A Minimum Risk Trigger Index, *Nucl. Mater. Manage.*, 8(2): 98-100.
- Wald, A., 1950, *Statistical Design Functions*, John Wiley & Sons, Inc., New York.

ADDITIONAL READING

- Beetle, T. M., 1976, Quantifying Scenarios to Check Statistical Procedures, *Nucl. Mater. Manage.*, 5(3): 612-621.
- Bennett, C. A., 1973, Some Effects of False Positives, *Nucl. Mater. Manage.*, 2(3): 110-125.
- , 1980, Data for Detection and Analysis, *Proc. 2nd Annual Symp. on Safeguards and Nuclear Material Management*, March 26-28, 1980, European Safeguards Research and Development Association, 11: 185-187.
- Markin, J. F., A. L. Baker, and J. P. Shipley, 1980, Implementing Advanced Data Analysis Techniques in Near-Real-Time Materials Accounting, *Nucl. Mater. Manage.* (Proc. Issue), 9: 236-244.
- Schlefer, J. C., 1975, A Decision Structure for Plant Safeguards, *Nucl. Mater. Manage.*, 4(3): 406-430.
- Shipley, J. P., 1979, Data Analysis for Nuclear Material Accounting, *Proc. 1st Annual Symp. on Safeguards and Nuclear Material Management*, April 25-27, 1979, European Safeguards Research and Development Association, 10: 391-395.

CHAPTER 13

Nuclear Material Accounting

13.0 INTRODUCTION

The preceding chapters presented the fundamentals of probability and the common statistical methods of processing and analyzing sets of numerical data. The remainder of this book is devoted to the application of these methods to nuclear material accounting. This chapter explains material accounting practices and the use of statistics in both controlling the measurement results needed for nuclear material accounting and evaluating the results of the accounting process.

For the benefit of readers without material processing or material accounting experience, the appendices to this chapter provide information on the structure of a typical nuclear material accounting system and a brief description of the most common measurement procedures. This information is presented at a very elementary level, but it is intended for readers without training or experience in these areas. It will help provide a background that is important for understanding the models used in subsequent chapters to define procedures for estimation and evaluation of material accounting results.

13.1 THE MATERIAL ACCOUNTING PROCESS

13.1.1 Basic Material Accounting Practices

Maintaining a record of possessions is a universal practice. All acquisitions are added to the inventory records and all removals, sales, losses, discards, and other disbursements subtracted. Manufacturing and mercantile establishments maintain such inventory records to have a current **book inventory**. An accurate book inventory is needed for many activities:

1. Planning and scheduling production.
2. Scheduling supply orders.
3. Making sales commitments.
4. Calculating monthly profits and net worth.
5. Complying with legal obligations, such as reporting profits and net worth.

Periodically, business establishments also perform a **physical inventory**; i.e., they make a complete check of all material on hand. The purpose is to check the accuracy of the book inventory and make corrections as appropriate.

Maintaining a book inventory of materials and reconciling it to a physical inventory is the essential principle of all material balance accounting. In the nuclear material area, material balance accounting provides tests of the effectiveness of the other material control methods as well as providing for periodic adjustments of the records. The other material control methods are physical protection of the facility and material and personnel control and surveillance. Typical methods of control are personnel access controls, assigned custodianship of materials, monitoring of material and personnel movements, and the use of alarmed storage areas and vaults.

Normally, book and physical inventories differ because of the accumulated effects of unrecorded breakage, spoilage, pilferage, leaks, and other losses, as well as bookkeeping mistakes and omissions in recording additions and removals or taking the inventory. Inventories of bulk materials such as fuels and grain are also subject to errors in the measurements of the quantities on hand. Consequently a book-physical inventory difference, or simply an **inventory difference**, is observed. On the basis of experience or the value of the material, it may be possible to judge whether the inventory difference is so large that it merits an investigation into the causes. After significant discrepancies have been resolved, the book inventory is adjusted to correspond to the physical inventory. The inventory accounting of firms subject to federal regulatory agencies, such as the Securities and Exchange Commission, must be reviewed annually by independent auditors. This normally includes witnessing by auditors of the physical inventory-taking process.

The model which describes this closed system of input-output accounting over a fixed time is

$$ID = BI + A - R - EI \quad (13.1)$$

where

ID = inventory difference†

BI = beginning physical inventory

A = additions to the inventory

R = removals from the inventory, i.e., discards, known losses, and shipments of products and by-products

EI = ending physical inventory

If the quantities on the right of the equation represent a complete and exact determination of quantities added, removed, and in inventory, the inventory difference will be zero. As noted previously, observed values of

†The term "Material Unaccounted For" (MUF) is also commonly used, especially in international safeguards, to express this difference between net transfers and inventory change. In some usages, MUF denotes specifically the expected, rather than the observed, inventory difference.

the inventory difference will usually differ from zero because of accounting errors, measurement errors, and unknown or unmeasured losses. This distinction between the observed and expected values of the inventory difference, defined by Equation 13.1, is essential to proper application of statistical procedures to the observed differences.

Material accounting systems are similar to financial accounting systems. Financial accounting involves recording the information concerning the economic resources of the firm, which are reflected in its assets, liabilities, receipts, and disbursements. Inventory accounting is concerned with the material resources of the firm as expressed by material flows and inventory. The following are key elements of both accounting systems:

1. A unit record (source document) for every transaction.
2. Records of transactions and inventory status in the form of journals and ledgers to record, classify, and summarize the data.
3. Double-entry procedures to provide internal control and help prevent errors in the records.

Financial accounting systems have account categories such as assets, liabilities, and sales. Material accounting systems have analogous account categories such as material type, material location, shipments, and receipts. These accounts may include financial value information as well as the quantities of material.

Computerized accounting systems provide a great deal of information-handling capacity and flexibility that can be used to record supporting information as well as basic accounting data. Many inventory accounting systems are essentially real-time systems in that the information about the material is updated almost immediately after completion of each transaction affecting the material status. Such systems are generally accessed directly from remote, manually operated terminals and, in some cases, automatically by instruments that identify and measure the quantities of material.

13.1.2 Nuclear Material Accounting

Nuclear material accounting retains most of the basic accounting practices characteristic of the systems generally used for property accounting. The unique aspects of accounting for nuclear material arise primarily because of complexities in handling and measurement due to the special character of nuclear material.

The accounting system required at a nuclear facility is not greatly affected by the nuclear characteristics of the material. Some characteristics of the accounting system, such as its size, whether it is computerized, and the types of documents used, depend on the type of facility, number and type of transactions to be handled, types of materials, the information needs of management, the internal controls needed, and legal requirements

imposed by regulatory agencies. The system should provide a continuous book inventory of the nuclear material possessed by the facility and generate reports of receipts, shipments, discards, losses, and material status as required.

In some cases, nuclear material can be accounted for as uniquely defined items, such as reactor fuel elements and other fixed units of nuclear material. When the accounting involves material processed in bulk form, however, transfers and inventories must be based on the quantities in batches of this bulk material. The measurements of those batch quantities are an important element of nuclear material accounting. The measurements are usually supported by a measurement control system that serves both the nuclear material accounting system and the production operations. It provides the information needed to assure the quality and validity of the nuclear material accounting data. Through the measurement control system, the loss detection sensitivity of the material balance system is improved, measurement and product quality are enhanced and, most importantly, figures of merit can be established for both the products of the facility and the nuclear material control and accounting system.

A frequent consequence of the difficulty of direct measurement of certain nuclear material flows, such as wastes and process effluents, is that the material balance concept may sometimes be used to provide a **by difference** value for some of the quantities entering into the material balance model. This means of accounting should be avoided because it leads to a less meaningful inventory difference estimate. If the quantity of nuclear material remaining in the inventory at any location in the facility, or that contained in a discard, is determined by difference, the capability of the material accounting system to detect a loss or a gross measurement error at the location in the facility can be substantially reduced or even destroyed.

13.2 IMPLEMENTATION OF NUCLEAR MATERIAL ACCOUNTING

The effective application of the concepts of material accounting to nuclear material requires an understanding of the function, design, process characteristics, and material flows of the specific facility; the nature of the accounting systems available to describe these facilities; and the measurement systems required to determine the nuclear material content of the material flows and inventories. Correct application of statistical methods in connection with nuclear material accounting requires familiarity with the facility and process characteristics, accounting principles, and measurement problems that are basic to the implementation of nuclear material accounting at a given facility.

13.2.1 Facility Design and Process Flow

The types of facilities with sufficient special nuclear material (SNM) to require a formal material accounting system range from research laboratories to uranium refineries and processing plants for spent nuclear fuel. Table 13.1 gives a list of the common types of facilities to which nuclear material accounting procedures are applied and the functions and accountability characteristics of each. Commonly, two or more of these facilities are incorporated into a single production plant or complex.

TABLE 13.1
Common Types of Facilities Subject to Formal
Nuclear Material Accounting Requirements

| Type of Facility | Function | Characteristics |
|--------------------------|---|---|
| Uranium refinery | Purify uranium ore concentrates | Natural uranium, only; a wet chemical process with high flow to inventory ratio; large liquid waste volumes and little scrap recycle |
| Uranium conversion | Convert uranium oxides to hexafluoride (UF ₆) | May process uranium of several enrichments including natural uranium; a chemical process of primarily gas phase methods; high flow to inventory ratio |
| Uranium enrichment | Isotopically enrich UF ₆ | Wide range of uranium enrichments; gas phase, continuous processing; large in-process inventory; low waste volumes; considerable solid scrap to reprocess |
| Nuclear fission reactor | Generate neutrons for research or thermal energy for commercial use | All special nuclear material (SNM) in encapsulated item forms; generally no waste or scrap in non-item form |
| Fuel material conversion | Convert UF ₆ to fuel grade oxide (UO ₂) | Chemical processes, both wet and dry, may be batch or continuous flow process, or a combination; may process several enrichments; moderately high inventory/flow ratio |
| Uranium-fuel fabrication | Fabricate reactor fuel, either low, intermediate or high enrichment | Product may be pellets, rods or pins, plates or assemblies. Generally batch processing. Dry powder and mechanical/metallurgical processing; moderate inventory/flow ratio; small waste quantities; large scrap quantities |

(Table continued on next page.)

TABLE 13.1 (cont'd)

| Type of Facility | Function | Characteristics |
|------------------------------------|---|--|
| Uranium-plutonium fuel fabrication | Fabricate UO_2 - PuO_2 (mixed oxide) reactor fuel | Dry powder and mechanical-metallurgical processing or wet chemical processing; generally batch processing; moderate inventory/flow ratio; all processing in glove-box or remote operation enclosures; large scrap recycling; small quantities of waste; product may be pellets, rods or pins, plates or assemblies |
| Chemical reprocessing | Recover uranium and plutonium (optional) from spent fission reactor fuel | May be high, intermediate, low, or depleted enrichment; no segregation of enrichments within those classes; primarily wet chemical processing; continuous or batch/continuous combination; products are oxides of uranium and plutonium; most of process is remotely operated; large volumes of radioactive waste; moderate scrap quantities either moderate or low inventory/flow ratio |
| Research laboratory | Develop basic knowledge or methods of processing or fabrication of fuel materials, fuel material refining, fuel fabrication or recovery | May be wet chemical or metallurgical and ceramic experimentation on a small scale; large inventory to flow ratio; may be several kinds and forms of nuclear material and batch operations; small waste quantities and relatively large scrap quantities |
| Pilot plant operations | Test nuclear process operations on a semiplant scale | May be chemical, metallurgical, or ceramic process or some combination; generally large inventory to flow ratio; intermittent operations, either batch or continuous; comparatively large scrap and waste generation |
| Critical facility operations | Test nuclear reactivity of critical assemblies; develop basic criticality data | Generally only sealed items rather than bulk SNM handling unless the facility also fabricates specimens and assemblies for test; high inventory to flow ratio; may combine features of research and critical facilities |
| Scrap processing | Recover and purify nuclear material from production and research scrap | Similar to chemical reprocessing except that fission products and high radiation levels are not involved |

The following information is necessary for establishing an accounting system:

1. A diagram of the nuclear facility and its operations (for a manufacturing facility, a process flow diagram is recommended).
2. The types of nuclear material and typical quantities in the inventory and flows at the key operating points.
3. Typical batch and lot sizes.

This information forms the basis for determining key elements of the accounting system:

1. The logical subdivisions of the facility for material balance accounting, such as material balance areas (MBA) and item control areas (ICA).†
2. The key measurement points and the measurement methods used.
3. The sources of uncertainty affecting the observed inventory differences.

The facility characteristics given in Table 13.1 are some of those having an appreciable effect on the accounting system. The enrichment of the uranium is mentioned because the accounting and reporting requirements for high-enrichment materials are generally more stringent than those for lower enrichments. In addition, material of different enrichments is separated in the accounts as well as in the process (except that limited commingling may be accepted in scrap recovery). The processes described in Table 13.1 are also identified as batch or continuous because their accounting data collection methods usually differ. For example, batch processes generally result in many more items for bulk measurements than continuous processes. In addition, samples and laboratory analyses for batch operations are traceable to predictable and relatively constant quantities, namely production batches and lots, whereas material sampled from continuous processes is usually associated with time intervals rather than batches of material.

A simplified diagram of a nuclear material manufacturing process is shown in Figure 13.1 to illustrate the type of process information necessary to design an accounting system. The principal processing steps are blocked out, and the main nuclear materials are shown on the flow lines. The key measurement points are also shown.

The two storage areas shown, Cylinder Storage and Scrap Storage, are usually regarded as individual ICAs, and their inventories are accounted for separately. The receiving and shipping area would be an MBA in

†The term "item control area" (ICA) is applied when only uniquely identifiable items, such as sealed or tamper-safed containers whose contents have been previously determined by measurement of SNM, are stored in the area, and accounting by item identification only is adequate. The significance of MBAs and ICAs is that all inputs and outputs are measured, and book inventories are maintained of each MBA and ICA.

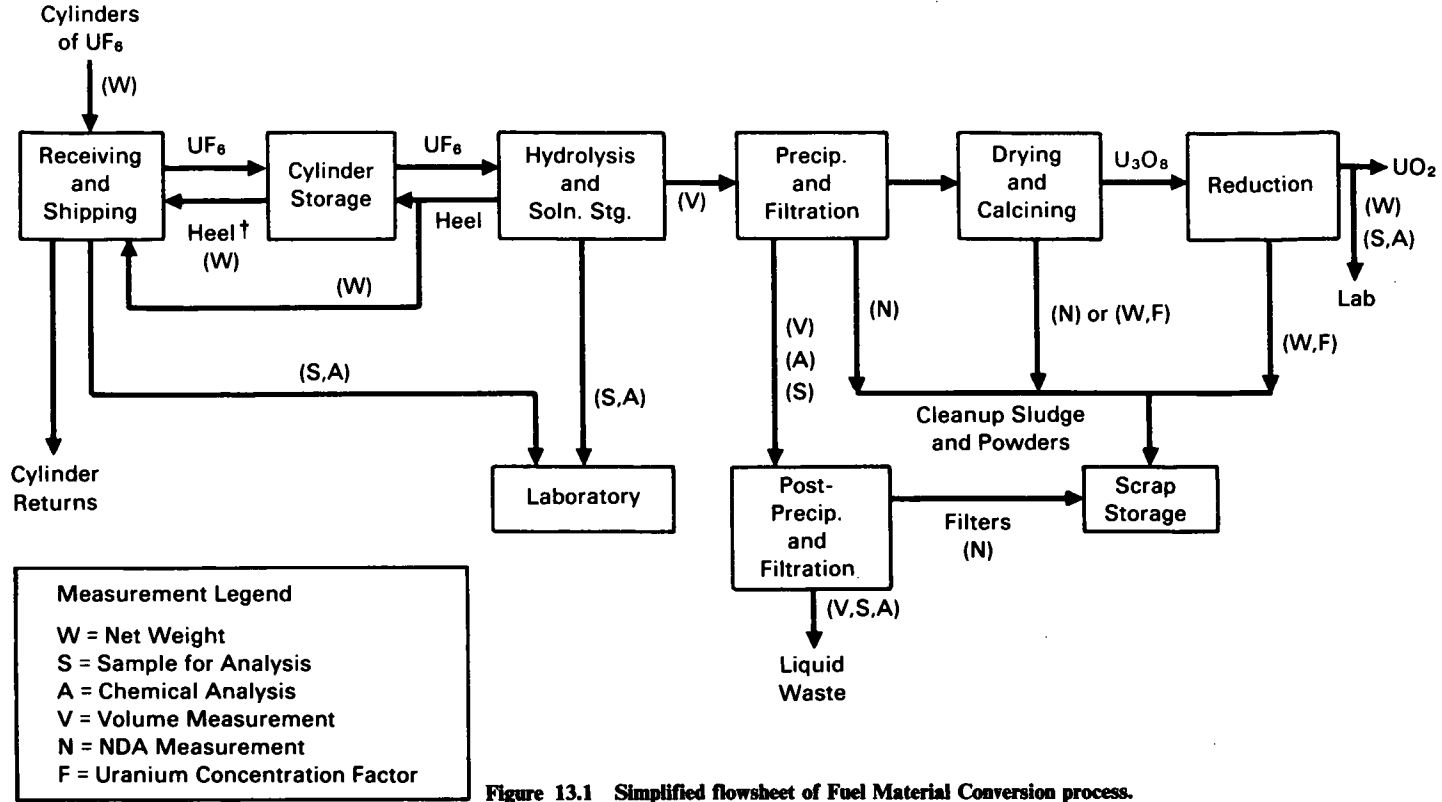


Figure 13.1 Simplified flowsheet of Fuel Material Conversion process.

†The term “heel” refers to the residual UF_6 in spent cylinders. Not all of the contents are removed in the hydrolysis process, and the cylinder containing residual UF_6 is weighed before it is returned to the UF_6 supplier for refilling. The net UF_6 into the process, therefore, is the full cylinder weight minus the returned cylinder weight.

which the containers received and shipped are checked by gross weight, seal, and container identification and samples are taken for laboratory analyses. The remainder of the process diagrammed in Figure 13.1 may constitute just one MBA in which bulk materials are processed. It is bounded by the UF_6 inputs to the hydrolysis step and the following outputs:

1. UF_6 heel returns.
2. Samples to the laboratory.
3. Liquid waste.
4. Filters and solid scrap to scrap storage.
5. UO_2 powder product.
6. Waste removals from work areas (contaminated tools, gloves, plastic, and paper) and airborne powders from the drying, calcining, and reduction areas (the waste removals are not shown in this diagram).

13.2.2 The Accounting System

In the United States, material accounting requirements are specified by three agencies:

1. The Nuclear Regulatory Commission for facilities, individuals, or organizations licensed under Title 10, Part 70, of the Code of Federal Regulations.
2. The Department of Energy for its own facilities and the contractors to the Department of Energy.
3. The Department of Defense for its own facilities.

The accounting practices and measurement and control procedures that prevail in the nuclear industry reflect the requirements of these federal agencies.

Source data forms provide the mechanism for entering transactions data into a journal and, subsequently, into a ledger. The following source data forms for nuclear material accounting are typical:

1. Nuclear Material Transaction Report (DOE/NRC Form 741).
2. Receiving Report.
3. Shipping Report.
4. Waste Discards (a shipping report may be used).
5. Internal Transfers.
6. Project Transfers.
7. Physical Inventory Report.
8. Journal Entry.
9. Analytical Request and Report Form.

All of these forms except the first are generally created at each facility to satisfy its particular needs. They are usually prenumbered sequentially to facilitate record keeping and the traceability of information through the accounting system.

The ledger consists of the inventory, ownership, project, material type, receipts and shipments accounts, and possibly others. Common transactions that affect the ledger balances are receipts, shipments, discards, transfers, and adjustments to the inventory. The adjustments category may include changes for the following reasons:

1. Inventory differences.
2. Shipper–receiver differences.
3. Measurement and bookkeeping errors.
4. Bias corrections.
5. Radioactive decay and fission or transmutation.
6. Change of enrichment, such as by commingling of materials.
7. Change of ownership or project number.

Facilities using an electronic data processing system can achieve nearly real-time accounting but a time lag between completion of a transaction and the updating of the accounts may be several days to 1 or 2 weeks, depending primarily on how data are entered into the system and how frequently the facility management chooses to update the data base. Another reason for time lags is that element and isotopic composition measurements by chemical and mass spectrometry methods usually require a day or two before the results are obtained. The time lapse for isotopic analysis data may be more than a week if the analyses are performed off site.

Internal controls for the bookkeeping system should be maintained to minimize mistakes and omissions. All changes to the accounts must be based on authorized and signed paperwork. To help detect mistakes, the accounting procedures may provide that no one person handles a transaction from beginning to end. When transfers are made that involve changes in custodial responsibility or changes in location, two individuals should confirm the accuracy of the transaction. These are generally the person initiating the transaction and the person receiving the material.

Referring to Figure 13.1, note that the following material movements generate accounting source data:

1. Receipt of a cylinder of UF_6 . (This generates a receiving report. The identity and gross weight are recorded and compared to the shipper data on the Form 741 to verify the correctness of the transaction record. Analytical data are confirmed later.)
2. Return of an empty UF_6 cylinder with heel to the supplier. (This generates a shipping report.)
3. Sample sent to the laboratory. (The laboratory maintains a **sample log** or subsidiary ledger to account for its inventory of samples.)
4. A UF_6 cylinder is transferred to cylinder storage. (Some facilities may keep a **storage log**, and the centralized accounting system may not keep records distinguishing UF_6 receipts from UF_6 storage.)
5. A UF_6 cylinder is transferred to the process. (An Internal Transfer Form is used.)

6. An "empty" cylinder is transferred from the process to storage.
7. A container of product (UO_2) is transferred from the process to a storage area (ICA) or another process.
8. Scrap is transferred to scrap storage.
9. Liquid waste is transferred to a treatment or disposal facility.†

13.2.3 Measurement Systems

The accounting entries that permit the determination of the quantities of nuclear material in shipments, receipts, discards, and inventories are obtained by measurement. Numerous measurement methods are used, and they are applied to a large variety of material types. The methods fall into four categories: bulk measurements, nondestructive assays (NDA), chemical analyses, and isotopic analysis. Material sampling must also be included in any discussion of the measurement system because it affects the errors in chemical and isotopic analyses, even though it is not itself a measurement method.

The nuclear materials measured are likely to include relatively pure compounds and alloys of the uranium and/or plutonium, production intermediates (consisting of at least some mixtures and low-purity forms of the elements), production scrap, wastes in solid and liquid forms, solutions covering a wide nuclear material concentration and purity range, and encapsulated items such as fuel rods, plates, and assemblies. Appendix 13B contains a brief description of the common measurement systems used and their characteristics.

The terms **item**, **batch**, and **stratum** as used in nuclear material accounting have specific meanings that may not be the same as the common understanding of the terms among engineers and production managers. An **item** is an object or container whose content can be uniquely determined and accounted for or transferred as an entity. A vessel or tank containing nuclear material in a solution can be designated as an item.

A **batch** is a quantity of material with a common element concentration or element factor. A batch may consist of many items, each having a unique weight or volume associated with it.

The term **lot** is commonly used in manufacturing for a quantity of material that has had uniform process treatment and was produced as a single entity. It may have a uniform composition identified by a single concentration factor. The term **lot**, however, is not specifically used in nuclear material accounting.

†Liquid wastes may be handled in several ways: (1) accumulation in holding tanks and transfers by batches, (2) continuous transfer to holding tanks in another area and analysis by batches, and (3) continuous transfers to another area with continuous monitoring of flow volume and uranium concentration. In the latter case a periodic Internal Transfer or Waste Discards form will be prepared at designated intervals, such as daily, to record the total transactions for that period.

The term **stratum** is applied to a collection of batches of like material that are in the same **component** of the material balance equation, i.e., the beginning inventory, receipts, removals, or ending inventory. Examples of strata are batches of UO_2 powder in beginning inventory, a group of similar cylinders of UF_6 in receipts, and a set of containers of UO_2 grinding powder in ending inventory. Within a stratum the number of items per batch should be similar, and the procedures for weighing or volume measurement and the sampling and measurement systems for determining chemical and isotopic concentration should be identical. These conditions are important to the variance calculations necessary for planning inspections and evaluating inventory differences.

13.3 THE ROLE OF STATISTICS

Statistical methods have been especially useful in determining the effectiveness of nuclear material accounting because of the unique problems of material handling and measurement. The initial use of statistical techniques was to evaluate the uncertainty in the reported amount of nuclear material received, produced, shipped, lost, discarded, or otherwise added to or removed from inventory. Statistical methods were developed to estimate and control measurement error and other sources of uncertainty in the data which are the input to the accounting process and to determine the consequent uncertainty in the accounting results. More recently, statisticians have extended their concerns to the effect of these uncertainties on the inferences and decisions possible from the accounting results. The basic issues are the power of the accounting system to detect process anomalies and, conversely, the assurance provided by the accounting results that such anomalies are absent and the accounts are correct.

There is a strong analogy between the role of statistics in material accounting and its role in determining the reliability and/or effectiveness of any complex system. Methods are required for controlling and estimating "component" reliability and also for combining these into a measure of "system" reliability that appropriately reflects the system's effectiveness. The approaches used and the problems encountered are those which normally arise when statistics are used to evaluate performance or establish optimal decision processes.

The following sections present an overview of the statistical approach to the analysis and evaluation of nuclear material accounting results. The discussion is in three parts:

1. Modeling the effect of measurement errors on accounting results and deriving formulas for variance estimation.
2. Estimating and controlling measurement uncertainty.
3. Estimating losses and formulating decision criteria using observed inventory differences.

The ultimate concerns of all the statistical procedures are the power of the accounting system to detect anomalies and, in the absence of anomalies, to provide assurance that the nuclear material has been accounted for adequately.

13.3.1 The Effect of Measurement Uncertainty

An observed inventory difference is a complex combination of thousands, or even tens of thousands, of individual measurements, each of which may contain one or more measurement errors. To determine the variance in an observed inventory difference from the variances of the error effects present, it is necessary to model the accounting process in statistical terms. This model is the connection between the inherent capability of the accounting and measurement system and the adequacy of the decisions based on the accounting results. Chapter 14 deals with the detailed procedures for describing the many possible error variances and determining their contribution to the variance in the observed inventory difference.

The modeling process starts with the simple expression for an inventory difference in terms of receipts, removals, and beginning and ending inventories given by Equation 13.1. For some purposes it is convenient to consider receipts minus removals as a single component called **net transfer** or **inventory change**. Following the terminology of Section 13.2.3, at the next level each of these components is considered to consist of one or more strata containing similar material, which implies the same measurement procedures and hence comparable measurement error variances. Strata, in turn, consist of batches and items, which, for modeling purposes, are determined by the fact that they are associated with unique random measurement effects. These unique effects may still, however, be a combination of effects from several primary sources. For example, one stratum of an ending inventory may consist of 100 containers of UO_2 powder divided into two batches of 50 items each. The uranium content of each container is determined by multiplying its individual recorded weight by a uranium concentration for each batch. The weight contains a random error effect combining errors in gross and tare weighings, and the uranium concentration contains a single random error effect combining the sampling and analytical errors for five random samples from the 50 containers.

Once the individual quantities that contribute to the observed inventory difference are known, each can be expressed in terms of its expected value and associated random error effects. Then, the associated variances can be combined on the basis of the error effects present in each level. This process of expressing the inventory difference as a linear combination of error effects rather than observations accomplishes three purposes:

1. It ensures proper cancellation of the contribution of identical effects which appear with opposite signs.

2. It ensures proper recognition and treatment of the occurrence of identical random effects in more than one result.

3. It ensures the recognition and definition of a minimum number of basic random variables associated with the effects present.

The expression of the inventory difference in terms of independent error effects rather than correlated errors in item quantities greatly simplifies the application of the formulas of Chapter 2 for determining the properties of functions of several random variables to the calculation of the variance of the inventory difference.

13.3.2 Estimating and Controlling Measurement Uncertainty

Chapter 15 gives detailed procedures for estimating error variances from measurement data. The application of statistics to a measurement system has two principal objectives. The first is to establish and maintain a state of control over the measurements, in the same sense that a quality control program is applied to identify and remove assignable causes of variability from a production process. This encompasses procedures for eliminating gross errors of measurement and known sources of bias. The second objective is to evaluate the inherent uncertainty of the measurement system. Statistical methods are required, both to provide assurance that the measurement system is performing as expected and to estimate the error variances necessary for the evaluation of system performance.

13.3.2.1 Measurement Control Programs

International guidelines for State Systems of Accounting and Control recommend that a measurement control program be capable of “ensuring, inter alia, that the adequacy of routine operation of the measurement systems is confirmed; that measurement systems are recalibrated at appropriate intervals; that random and systematic errors† are properly estimated for propagation so that the limits of measurement uncertainties associated with MUF can be established; and that clerical errors are, so far as practicable, detected and corrected.” (IAEA, 1980).

Measurement control programs are frequently focused on the process of measurement rather than the data produced. From the point of view of the data production process, this is probably a correct emphasis. There is an old quality control adage that says, “You can’t inspect quality into the product.” From the view of the user of the data, however, operational errors arising from deficiencies in accounting and in recording and reporting data are just as important as measurement errors. Operational errors that influence data quality include deficiencies in procedures, such as the

† The terminology in this quotation differs from that used in the subsequent chapters of this book (Section 14.2.1).

existence of unmeasured losses, unmeasured inventory, and mistakes in recording due to misplaced or miscounted items. Other operational errors include mistakes in the computation, tabulation, and reporting of the information, and errors due to accounting deficiencies such as incomplete material descriptions or the incomplete or erroneous definition of MBAs. The nature of the errors arising from these operational and accounting deficiencies can range from gross errors or mistakes to biases resulting from longstanding procedural deficiencies, such as might arise from the existence of unmeasured waste losses.

Data quality must continuously be monitored at key control points and periodic audits of the complete system must be provided. Statistical sampling and control techniques are important for designing such a system. All source documents, ledgers, journals, and reports in the material accounting records system need to be monitored for errors in calculation and transcription. Tests need to be designed into the accounting system that check the completeness, consistency, and accuracy of the data received. No system can guarantee that no mistakes exist or that all mistakes are found and corrected. The system should be designed, however, to control the magnitude and frequency of the errors so that the cumulative effect is inconsequential to the results obtained. This is possible only if the system is designed to provide timely identification of the reason for lack of control rather than mere indication (after the fact) that a problem exists. Procedures are necessary to ensure that effects on the material balance due to accounting errors are identified and corrected.

Statistical techniques can contribute to the program for monitoring and controlling the measurement procedures as well as the accounting processes. A thorough measurement control program should monitor all sources of measurement error needed to estimate the measurement uncertainty in the accounting results. Sufficient measurement control data should be obtained during each material balance period to independently estimate the variance components for each source of measurement error or to confirm that existing estimates are representative of the current performance of the measurement system. Measurement control data collected during earlier periods should not be combined with current data if statistical tests indicate that the two sets of data came from different populations. It cannot be assumed that data collected in the past are applicable to current measurements. Frequent monitoring, distributed uniformly over the time that the measurements are made, tends to ensure that measurement data reflect current conditions.

A measurement control program must include calibration and standardization procedures designed to produce results that are free of bias. The two principal requirements for an effective program are the existence of adequate reference standards and the necessary procedures for ensuring that the measurement methods are continuously monitored against these

standards. The statistical methods used in connection with procedures for calibration, standardization, recalibration, and the determination of bias corrections or bias correction factors are derived from standard regression and hypothesis testing methods. There is an extensive literature available on the application of statistics to calibration problems.

The program for routine testing of a measurement process using reference standards will vary with the nature of the measurement and with the circumstances under which the measurement is made. For each measurement method, the checks should be frequent enough that errors in calibration or changes in bias correction factors will not become a significant part of the measurement uncertainty. Some measurement systems must be checked daily. Others may be very stable and need much less frequent recalibration, as is usually true for scales and volume measurement tanks. The frequencies required can be determined from an analysis of historical information on calibration stability and the sensitivity of the material balances to biases in that particular measurement process and to the variance of the calibration. Calibration of a measurement tank or an instrument not readily accessible for recalibration should be sufficiently thorough that the variance of the calibration will be small compared to the uncertainty in the entire measurement process. In high throughput processes, calibration variances can dominate the total measurement variance of a material balance, so that the optimization of the standardization and calibration procedures for the measurements involved can become an important element in the control process. The maximum desired calibration variance for any particular application should be estimated to provide design guidance for the calibration. This can usually be done by making a statistical projection of the effect of the calibration variance on the total variance of the ultimate accounting result, i.e., a typical material balance or typical quantity of nuclear material in a shipment.

Effective calibration, particularly in the case of NDA devices, may require in situ experiments or special destructive measurements to establish its validity. This is particularly true in the case of equipment used to monitor the nuclear content of heterogeneous wastes or to make in-line measurements of liquid waste streams. Some general guidelines for control of measurement biases through effective calibration and bias monitoring have been published by Brouns, Merrill, and Roberts (1980). That publication also provides references to other literature on recommended calibration methods and sources of standard reference materials for calibrating the common measurement systems used in nuclear material accounting. The common sources of measurement bias for several of the measurement systems are discussed in Appendix 13B.

13.3.2.2 Estimating Measurement Variability

Measurement errors for nuclear material can occur in either the determination of bulk amounts or constituent concentrations. The easiest and

most precise bulk measurements are the mass determinations. The most difficult, and hence generally least precise, bulk measurements are the volume determinations, particularly when they must be made in irregular tanks or are associated with a continuous flow such as a waste stream. Analytical determinations of concentration are most precisely made on high-purity material that is relatively homogeneous. For relatively heterogeneous materials, such as many wastes, nondestructive methods which determine the total isotope or element content may be preferable to the use of elaborate schemes to obtain a sufficiently reproducible sample for precise analysis, even though an NDA method is generally less precise than a destructive analysis method.

It is important to recognize that from a particular set of data used to compute one or more safeguards indices, such as an inventory estimate or an inventory difference, it may not be possible to estimate, either individually or collectively, all of the individual contributions of the measurement errors to the estimated value of an index and its variability. A situation may arise in which the presence of an error source in the data is known, but estimation of its contribution is impossible from the data used to compute the index. This, for example, may be the case when it is known that a constant effect of unknown magnitude and direction is present because established values of quantities such as disintegration rates or standard concentrations of reagents are used in determining all results. This possibility may be accounted for by associating a variance component with the single fixed effect, even though no specific confirmation of the contribution to the total variance of the index is possible from current accounting or measurement control data. On the other hand, there may be unknown contributions to the error that have not been identified and, therefore, cannot be estimated or controlled statistically either from current accounting data or through a designed measurement control program. The possible contribution of such effects must be determined from technical considerations.

All of the statistical techniques used to estimate variance components associated with measurement error have the common feature that more than one independent measurement is required. The replicate data used to estimate these quantities should be independent and should include all routine procedural steps. All steps in the measurement procedure—including sampling, sample preparation, and analysis—should be replicated and should follow accepted procedures as closely as possible. It may even be necessary to separate the measurements in time or to have them made by different persons to ensure that all contributions to the measurement uncertainty are included in routine checks.

This necessary replication of measurement results can seldom be justified on the basis of the increased precision obtained by the replication, and therefore it must be recognized that the replicate analyses represent work done purely for the purposes of estimation and control. They allow an "inspection" of the quality of the analytical results produced. Most pro-

grams, however, correctly emphasize the maintenance of a controlled and effective measurement system rather than relying on an inspection process to ensure product quality. Although the latter can measure whether or not the measurement system is in a state of control, only the former can ensure that good measurements will be produced.

The variance components associated with errors of measurement can be estimated using data from the calibration procedures, standardizations, and methods testing necessary in establishing and controlling measurement procedures, as well as by independent replicate measurements done specifically for error estimation. The analysis of reference standards should be distributed throughout the time that the analytical method is being applied to ensure that the absence of bias is being continuously monitored or that estimated corrections are representative of all the results. Standards used for monitoring measurement biases should closely simulate process samples in form and composition so that the results are not affected by differences in composition between standards and process materials. Again, every effort should be made to perform all measurements of standards by procedures identical to those used for routine measurements.

13.3.3 Inferences from Material Accounting

Chapter 16 deals with the contribution of statistical methods to the inferences possible from nuclear material accounting results. These inferences are usually concerned with possible losses or diversions, or the assurance provided that loss or diversion did not occur. In the case of a single observed inventory difference and its associated estimated error variance, or a standard error of the estimate, methods and procedures from Chapter 4 for forming confidence intervals and testing hypotheses are directly applicable. Tests for cumulative inventory differences and other loss estimators based on more than one observed inventory difference are more complex because of the sequential nature of the testing procedure and the need to take into account the dependence between observed inventory differences, particularly consecutive inventory differences that depend on a common physical inventory.

The key statistical quantities associated with successive material balances are the initial inventory, the net transfers, and the ending inventories for each of n successive material balance periods. These $2n+1$ quantities are complex linear combinations of individual accounting entries, each of which may be a linear or multiplicative combination of several measured quantities. Dependencies may exist both because two or more of these quantities involve identical entries and because one or more entries have a common fixed error effect. An example of the first cause of dependency would be the use of an identical measurement of the content of a UF_6 cylinder as part of a net transfer and one or more subsequent ending inventories. An example of the second is the use of an NDA method with a

fixed calibration bias to determine accounting entries for both receipts and physical inventories. Error propagation problems arise in determining the contribution of the operational procedures and measurement processes to the variability in, and dependence between, the estimates of the physical inventories and net transfers and the resultant inventory differences.

Statistical methods are also effectively applied in the verification of accounting data. Domestic regulatory bodies are concerned with confirming the quality of the accounting and measurement systems. Additional problems arise in international safeguards where not only the quality but the validity of the data are subject to question. Chapter 17 presents statistical procedures for both types of problems, with particular emphasis on the use of sampling plans for both evaluation and validation. Validation is studied both as a two-stage process of confirming reported results and then using the validated results for loss estimation and as a single-stage estimation problem based on the combination of accounting results and data obtained through verification procedures. In either case, the problem is made more complex by the need to consider alternatives involving deliberate falsification of reported data.

APPENDIX 13A

The Accounting System

13A.1 INTRODUCTION

The material accounting system of a nuclear facility provides the management with timely and accurate information concerning the types and quantities of nuclear material received and shipped, discarded or lost during processing, and in the inventory at the facility. The records of the accounting system should identify the location of each item, object, or other defined quantity in the nuclear material inventory; the item and batch identifiers needed to provide traceability to production and measurement data; and other information needed for process and quality control, production scheduling, and other management purposes. The information maintained by the system should also permit identification of item error effects, numbers of measurement replications, and other error path data (Section 14.2.1). The accounting system achieves these things through a system of source data forms, transaction journals, ledgers, and administrative procedures for collecting and compiling the data. For a more extensive treatment of nuclear material accounting than given here, the reader may consult the NRC report by Korstad (1980).

As discussed in 13.1.2, the book inventory of a facility is maintained by a conventional inventory accounting system that depends on the reporting and documentation of every transaction that affects the quantity or location of any nuclear material on the inventory of the facility. This documentation is the primary source of the information needed by the individual responsible for nuclear material control. The information consists of at least the following for each nuclear material receipt, transfer, shipment, discard, loss, and inventory change:

1. Quantity of the element and isotope (if applicable).
2. Identities of the containers and batches.
3. Material type.
4. Origin and destination, if a change of location is occurring.

To evaluate these data the following additional information is needed:

1. The basis of the quantity values, such as the measurement methods and specific instruments (scales, NDA instruments, etc.) used, and the dates of the measurements.
2. The sampling and analytical patterns, i.e., number of samples per batch and number of analyses per sample.

3. The quality control data applicable to the measurement systems used.

The sources of the accounting and statistical data are described for each type of change in material status in this appendix.

13A.2 RECEIPTS

When nuclear material is received by a facility, the receiver (individual or organization) is responsible for preparing a Receiving Report for the person responsible for nuclear material control. This report is normally prepared immediately with whatever information is available and amended reports are prepared later if necessary. Such reports describe the material, identify (usually by a numerical or alphanumeric code) the containers and batches, give the quantities in each item and batch, and note the current location of the material.

The receiver's quantity information is often reported to the accounting office in two or three steps. The bulk measurements can usually be made immediately upon receipt and reported at that time to the person responsible for nuclear material control, but the analysis results will be obtained a day or more later. When the element and isotope concentration values have been obtained, an amended receiving report should be issued.

The initial receiving report may give the quantity values obtained from the shipper. An amended report issued later will add the receiver's independently measured values. Some facilities accept all shipper's values for which the shipper-receiver difference (SRD) is not statistically significant. The receiver's measurements are made to verify the values and detect any significant SRDs that must be investigated and resolved. Other facilities record all receipts at their own values and record the differences between them and the shipper values in the SRD account in the general ledger.

Other statistical information needed for an analysis and evaluation of the receipts data may be obtained in various ways. The Analytical Request and Report Form generally gives the number of samples and what items and batches they represent, and it may give the number of analyses performed per sample, the methods, the analyst, the instruments used, and the dates of the analyses. An alternative source of this information in many facilities is a manual of standard procedures specifying sets of measurement methods for the facility for each category of material routinely flowing through the process. Unless exceptions are indicated, it is assumed that the standard procedure is followed. The procedures specify the weighing procedure; sampling method; analytical method; and number of weighings, samples, and analyses to be obtained.

The data from which the statistician can obtain bias and variance estimates for the receiver measurements are usually not available from the

measurement data itself because insufficient replicate measurements are made on individual items received. The statistician normally estimates the measurement biases and variances applicable to a receipt from measurement control program data.

13A.3 INTERNAL TRANSFERS

An internal transfer report is issued by the custodian of an item or batch of nuclear material when the material is transferred to a different custodian or to a different material balance or item control area. The report will identify the item and batch numbers, type and quantity of material, and the locations and custodians involved.

Usually the statistician does not need any information about the measurements of the nuclear material in internal transfers unless the measurement results will be used for a shipment or an inventory report. Most internal transfers involve movements of material between processing areas or from a process to temporary storage.

13A.4 SHIPMENTS

When a shipment is planned, a Shipping Report is prepared that describes the material and quantities in sufficient detail for the accounting office. The statistician may estimate the variance and the bias correction, if any, for each item or batch in the shipment for the Nuclear Material Transaction Report (DOE/NRC Form 741) on an individual shipment basis. The variances are commonly calculated from standardized or average statistical parameters that have been preestablished for each type of routine shipment. More detailed and specific statistical data, however, may be needed for the variance estimates of some shipments for incorporation into the variance of an inventory difference or to evaluate a shipper-receiver difference. For these purposes, the variances of the element factors, isotope factors, and masses on an item and batch basis will be desired. If a standard procedure was not specified, as discussed in Section 13A.2, an Internal Transfer Report may provide such information as the scale or NDA instrument number, date, and sample number, if any, and the Analytical Report will provide such details for the analyses. In this way the measurement information for the transaction is traceable, and the statistician can calculate the variances when they are needed in the preparation of a Shipping Report.

13A.5 WASTE DISCARDS

The nuclear material custodian who initiates the removal of items or batches of waste from the facility prepares an appropriate report, which will be referred to in this section as a "Waste Discards Report." This

report provides the identity, quantity, point of origin (process or material balance area), and disposition of the material. Solid waste—which may consist of paper, plastic, gloves, rags, clothing, contaminated metals and ceramic materials, and contaminated equipment—are generally packaged in plastic bags, cans, or cartons at the point of origin, such as the work station or hood. Then the packaged items are placed in 55-gal barrels or, in the case of large metal objects and equipment, in cartons or crates, and sealed in preparation for shipment to a disposal site. In most cases, the filled barrels and large containers, rather than the smaller individual packages, are analyzed for the nuclear material content. This is generally done by an NDA method based on gamma, neutron, or neutron coincidence counting.

If the Waste Discards Report does not provide sufficient information about the waste assay, the statistician will acquire this information from the assay report. Since the types of wastes and the assay procedures used are highly standardized in most facilities, the statistician will generally be able to determine from the material type description what assay method and, probably, what instrument was used. The measurement control program reports will provide the bias and variance monitoring data. The statistician should know, however, what instrument performance controls and techniques are being used. For example, an important concern in NDA measurements is whether the size, shape, container type, and composition of the contained material in the reference standards being used and in the unknown items being measured are nearly identical, because a mismatch of the standards and unknowns is a common cause of NDA measurement error. To help minimize such errors, wastes are usually segregated by type of material when packaged.

Liquid wastes—such as process filtrates, evaporator condensates, and wash water—may be collected in a holding tank and sampled for analysis before being discarded or routed to a treatment facility. On the other hand, they may be piped directly to a holding tank in a treatment facility through an on-line monitor that will alarm if the concentration of the nuclear material exceeds the upper limit for discarding. In any case, at some point a Waste Discards Report is prepared for accounting records by the responsible custodian or production supervisor to document the removal of the batch of waste. This report may be issued for each batch (probably each tankful) or for the accumulated discharges for a specified period, e.g., for each shift. The statistician needs to know how the batch quantity (mass or volume) was measured, how the batch was sampled, and how the samples were analyzed. When methods are standardized the principal concern will be simply what departures, if any, were made from the standard procedures.

Two common characteristics of nuclear waste disposition that cause accounting difficulties are large measurement uncertainties and loss of

traceability to their exact source because of consolidation of wastes. The assays of solid wastes may have a relative standard deviation up to 20 or 30% because of heterogeneity, fluctuating background radiation, and self-absorption or attenuation of the radiation within the container. The assays may also be biased because of significant differences between the matrix or distribution of the nuclear material in the unknown and the reference standards. The analyses of liquid wastes, which are generally done by chemical analysis, can usually be performed with a relative standard deviation of the order of 5%. The analyses can be biased, however, if the analytical methods are not adequately controlled or if the liquids contain nuclear material as suspended particles that are not detected in the analysis.

The traceability issue arises because wastes from more than one process area are commonly combined into one batch before the measurements are made. This is less common for solid wastes than liquids. When combined in this way the quantities of waste removed may need to be allocated between various material balance areas.

Another accounting problem associated with waste removals is the loss of material by unmeasured and, often, unknown paths. The loss mechanisms may be airborne material collecting in filters in the ventilation system, accumulation of material in hidden spaces of equipment and on the surface of ducts and work enclosures (which may be recovered eventually when the equipment is removed and work enclosures disassembled and cleaned or assayed), and material carried out on clothing to the laundry. It would seem that all loss mechanisms are or can be accounted for by measurements on all exhaust air filters, careful design of equipment and enclosures to minimize material traps and enable area NDA measurements of holdup, and routing of all sumps, flow drains, laundry waste and sinks, including those in clothing change rooms, to retention tanks for analysis before disposal. The practical situation, however, seems to be that small, unmeasured losses still occur.

13A.6 PHYSICAL INVENTORIES

A physical inventory is taken with an organized list of the items and batches of nuclear material on the book inventory. The items are checked off as they are located, and the quantities are recorded for each item or batch. The previously determined element and isotope quantity values given in the book inventory for items that are encapsulated or sealed with tamper-indicating seals are normally accepted without remeasurements. Other items are measured by NDA or by a bulk measurement plus a sample taken for analysis.

After completing searches for missing material, if any, and investigating any other discrepancies, the inventory difference is calculated and the

books adjusted to conform to the physical inventory. The total of the adjustments, which is the inventory difference, is entered into the ledger in an "ID Account."

To determine whether the inventory difference is statistically significant, the variances of the item and batch quantity values are needed. The basis for the variance estimates will usually be measurement control data for the measurement methods and instruments used. Frequently, the statistician can determine how many measurements were made, what methods were used, and what measurement control data are applicable by consulting the manual of standard procedures used for each type of material. When an exception to the standard procedure is made, however, the statistician must be aware of it and determine what measurement procedures were actually used in the physical inventory by checking the analytical reports and making personal contacts.

There are four common sources of accounting difficulty other than measurement uncertainty in conducting and reconciling the physical inventory:

1. **Material that is overlooked in taking the inventory.** It will be immediately known if an item listed in the records is overlooked, but that is not true if some of the bulk material that was to be packaged and measured for the inventory is overlooked.

2. **Misplaced material.** This can occur because of a failure to document a movement of material or because of a bookkeeping mistake.

3. **Holdup that is poorly measured.** Material held up in equipment, pipes, tank bottoms (heel), and filters may be so inaccessible that the measurements have a large potential error or that time-consuming clean-out for measurement is necessary. In some cases the inventory teams make an estimate of the quantity from mechanical or geometrical information or historical clean-out data.

4. **Holdup that is undetected.** Material is often contained or held up in locations unknown to or inaccessible to the inventory teams. The discrepancy becomes part of the inventory difference.

13A.7 INVENTORY ADJUSTMENTS

This category of the nuclear material transactions may include adjustments for inventory differences, SRDs, measurement and bookkeeping mistakes, bias corrections, radioactive decay, fission or transmutation, and changes of enrichment. The sources and flow of information for these adjustments are discussed individually.

13A.7.1 Inventory Differences and SRDs

After a physical inventory has been completed, the appropriate inventory accounts are adjusted to the physical inventory values and the inven-

tory difference is posted to the ID account. When an inventory difference is to be corrected after the end of the accounting period and an inventory account is also affected, the correction should be posted to the inventory account and a Prior Period Adjustments account instead of the current ID account, to avoid an error in the inventory difference estimate for the current period. Some facilities, however, record all adjustments in the current accounting period, maintain separate records of the effect on inventory difference, and apply a correction at the end of the accounting period.

SRDs are handled in a variety of ways. Some facilities accept and record the shipper's values for all received material unless the SRD is statistically significant. Other facilities record both shipments and receipts at their own values and record the SRDs in the SRD account. If an SRD is statistically significant, however, it is investigated and eventually resolved by a mutually acceptable procedure. Then the Shipments or Receipts account, whichever is affected, is adjusted to the accepted value unless the adjustment is made after the close of the accounting period, in which case the adjustments are posted to the SRD account and to a Prior Period Adjustments account. Some licensees do not use a Prior Period Adjustments account, however, but make corrections to the ID account at the end of the accounting period.

13A.7.2 Measurement and Bookkeeping Mistakes and Bias Corrections

Adjustments to the accounting records are necessary when mistakes and statistically significant biases are discovered after the material quantity data have been entered into the accounts. It is important that the adjustments result in a best estimate of the true material balance when the books are reconciled to a physical inventory. This may be accomplished in several ways.

Some facilities do not alter the existing accounting records but maintain an auxiliary record of all mistakes and biases. From that record, corrections can be made, if necessary, to the inventory difference to remove the cumulative effect of the errors. The auxiliary record can also be used to correct for errors in the book values of the quantities of material in shipments and other removals during the material balance period. A feature of this accumulation of error corrections is that some cancellation of error effects occurs and the final adjustments may be small.

An alternative way to handle each adjustment is to make corrections in every account affected by the error. This is rarely carried as far as correcting the records of each item or the tags on the items, however, because that would entail a great amount of paperwork. For example, hundreds of items are sometimes affected by one bias correction. A more common approach is to make the adjustment for the effect of an error as a single

entry in each account affected. That corrects the error in the inventory of the facility as well as in that account and avoids carrying it into the inventory difference at the close of the material balance period. If such an accounting correction has been made, it is not necessary to correct the book value of a batch of material as long as the material remains in the facility. The material balance is not affected by transferring material from one material balance area to another, although operators may correct book values when entering such a batch into a process to avoid errors in process yield or input-output data.

Unless book and tag values are corrected, auxiliary records of the error or bias in each batch value must be kept to help ensure that shipping reports and discard reports can be corrected. If a batch is split between shipments, it will also be necessary to allocate the error between the shipments.

13A.7.3 Radioactive Decay and Fission or Transmutation

Decay adjustments are calculated from nuclear constants whose bases and uncertainties are derived from the published literature. Fission and transmutation adjustments are based on reactor operating information (flux, exposure time, and neutron energy) and physical constants. Fortunately, the decay adjustment is a minor one for most facilities and is likely to have a minor impact on the variance of the inventory. The fission and transmutation adjustments apply only to nuclear reactor operations.

13A.7.4 Change of Enrichment

This adjustment is a shift of a material quantity between enrichment categories in the accounts and has a minor effect on the variances of the quantities. If the enrichment change is measured, the measurement variances are estimated in the usual manner from the backup measurement control data.

APPENDIX 13B

Measurement Systems

13B.1 INTRODUCTION

This appendix gives a brief and elementary overview of the measurement systems commonly used in the nuclear industry for material accounting. Most determinations of nuclear material content use measurements of the bulk amount of nuclear material and the concentration of the specific element and/or isotope desired. Bulk measurements may require the determination of the mass or volume of the material or the measurement of flows. Content can be determined by sampling and chemical analysis or directly by nondestructive assay (NDA). When more than one isotope of an element is present as in the case of enriched uranium, isotopic analysis may also be required. Each of these types of measurement is discussed in the following sections.

13B.2 MASS MEASUREMENTS

All mass measurement systems are based on comparison of the gravitational force acting on the unknown mass with the gravitational force acting on a known mass. The forces are measured by supporting the known or unknown masses on a force-sensing device or system. The net gravitational force is the measured force less the effect due to buoyancy.

Weight is defined by

$$W = mg \quad (13.2)$$

where m is mass and g is the acceleration due to gravity. The value for g varies with altitude, latitude, and local variations in the density of the earth. Weighing systems based on direct comparison with mass standards such as the mechanical beam balance are not affected by variations in gravity. Systems based on indirect comparisons such as load cells or electronic force-compensated scales that are calibrated in one place and used in another, however, may be biased by a significant amount. For example, the value of g is $981.624 \text{ cm s}^{-2}$ in Quiet Harbor, Alaska, and $979.053 \text{ cm s}^{-2}$ in Miami, Florida. This range could result in a difference of 2.6 parts per 1000. Calibrating the instrument at the place of usage removes this source of error.

Scales read directly in mass units, and calibration by the user consists of adjusting the instrument to read the exact value of a weight of known mass. Known masses (standard weights) that are "certified" to meet the accuracy tolerances recommended by the U. S. National Bureau of Standards are available commercially. Three common classes of "certified" standard weights are:

1. Class S, a "laboratory precision" class.
2. Class P, a "laboratory routine" class.
3. Class C, an "industrial routine" class.

The accuracy tolerances of these classes are about 2.4, 20, and 50 milligrams per kilogram, respectively. The accuracy of the standard weights, if protected from damage, will not significantly limit the accuracy of accountability scales as they are used in practice.

Most of the scales or balances used for nuclear material measurements fall in one of the following categories: †

1. Mechanical beam balances.
2. Load cell scales.
3. Electronic scales.

Mechanical beam balances are inherently very sensitive and can be designed to achieve high accuracy. This type of balance is available in capacities ranging from a few grams to thousands of kilograms and, therefore, can be used for nearly all SNM accounting applications. Load cell scales are compact and less subject to corrosion or damage by shock or vibration than mechanical beam balances. Remote operation is relatively simple, and they are readily adaptable for read-out at a remote location. These types of scales are now used most often for the mass range above 5 kg.

Common load cell scales in use by the industry are based on hydraulic, pneumatic, or strain-gauge sensors. A hydraulic scale measures weight by determining the hydraulic pressure generated when the unknown mass is placed on a piston in a hydraulic capsule. The pressure, which is proportional to the weight, is read out on a Bourdon tube or other pressure-indicating device. Typically, the unknown load is supported on three or more hydraulic capsules, and the net pressure is summed using a hydraulic totalizer.

The pneumatic scale is similar to the hydraulic scale. In this system, the load is supported on a diaphragm, and the weight is determined by measuring the air pressure required to support the unknown mass.

In strain-gauge load cells, weighing is done by measuring the change in length of a metal column or the deflection of a metal beam when sub-

† The terms "balance" and "scale" are commonly interchanged in usage and should be taken to have an identical meaning in this discussion.

jected to the weight of the unknown mass. Measurement of the change in length or deflection is performed by measuring a change in electrical resistance of strategically placed strain gauges on the column or beam.

Electronic scales are widely used in nuclear facilities for applications requiring capacities ranging up to about 50 kg. Like strain-gauge load cells, the output signal is electrical and, therefore, can be readily adapted to a digital reading system or direct input into a computer data base system. Also, the functions of taring† and zeroing can be automated. These features eliminate some of the sources of error related to human error and personal bias.

Some high precision balances used in laboratories are capable of weighing up to 100 g with a standard deviation of 0.05 mg or one part per 2,000,000. To realize this capability, however, the balance must be in excellent condition, and the operator must be skillful. Standard deviations of from one part in 20,000 to one part in 200,000 are available in rapid weighing, direct digital reading electronic and mechanical scales and balances. Multikilogram scales can be obtained with similar design features and standard deviation capabilities of one part in 20,000 or better. Routine weighings in nuclear processing areas, where most nuclear material weighings are made, however, are not usually so precise. Standard deviations greater than 0.1% relative are common. Some scales must be located in glove boxes or cells in processing lines where the environmental conditions are not ideal. The effects of dust or corrosive vapors will often increase the standard deviation of a scale considerably unless the scale's mechanical parts are frequently cleaned and repaired. Few, if any, of the higher-precision weighing devices are designed for such environments.

Since a scale may not be strictly linear, it should be calibrated over its range of intended usage. The accuracy of scales is monitored with reference standard weights to detect changes in performance. Bias of mechanical scales may be caused by corrosion or accumulations of dust on the pan or platform or moving parts; a nonlevel or unstable scale support, and damage to mechanisms such as the knife-edged pivots. Bias of load cell scales can be caused by hysteresis (which means that the response is affected for a short time after a change in load; i.e., return to equilibrium is delayed) and by a load cell creep phenomenon, as well as by the effects of dust and corrosion. Bias of electronic scales can be caused by electronic instability, mechanical damage, dust, and corrosion. Load cell and electronic scales, however, are not as vulnerable as mechanical scales to dust, corrosion, and mechanical effects like vibration and shock.

Statisticians should be aware of the potential sources of error in weighings. Some points to check are:

†Counterbalancing part of a load, such as that of an empty container, to return the scale reading to a set value such as zero.

1. Whether the scale has any drift, either long or short term, such as may be caused by temperature changes or accumulation of dust on the pan or working parts (drift sometimes leads to bias between gross and tare weighings and between input and output weighings for a process or MBA).

2. Whether the scale accuracy is affected by temperature changes or vibration or drafts that occur where it is used.

3. Whether the scale is moved often or subjected to occasional shock, such as when a heavy drum is set down nearby (these may change the level of the scale, shift the position of working parts, or damage parts, particularly knife-edge fulcrums).

13B.3 VOLUME MEASUREMENTS

Most measurements of the volume of a process liquid are based on measurement of the liquid level in a calibrated tank. There are several techniques: sight gauges, differential pressure gauges (manometers or pressure gauges), time domain reflectometry, conductivity probes, inductivity coils, resistance wire elements, ultrasonic gauges, and capacitive transducers. The most widely used are sight gauges and differential pressure gauges. Sight gauges and similar techniques, such as a graduated scale on a transparent tank (glass or plastic) or a transparent wall panel in a tank, are not subject to many sources of error. External sight glasses, however, can have a different liquid level than the tank if the liquid in the sight glass has a different density than that of the liquid in the tank. This occurs if either the temperatures or the solution composition in the tank and sight glass differ. The relationship between the densities and the liquid heights in the sight glass and tank is

$$h_1 = \frac{\rho_2 h_2}{\rho_1} \quad (13.3)$$

where

- h_1 = height of liquid in the tank
- h_2 = height of liquid in the sight glass
- ρ_1 = density of the liquid in the tank
- ρ_2 = density of the liquid in the sight glass

Differential pressure gauges (sometimes called pneumatic probes) are widely used for remote liquid level measurements, as is required with liquids in high radiation zones. The differential pressure may be measured with a liquid-filled manometer or an electronic pressure gauge (transducer). The liquid level is determined by measuring the difference in hydrostatic pressure between a point near the bottom of the tank and the space above the liquid. The equation relating the liquid height and the pressure difference is

$$\Delta p = gh_1\rho_1 \quad (13.4)$$

where

Δp = the differential pressure
 g = acceleration due to gravity

and h_1 and ρ_1 are as previously defined (Equation 13.3). If a liquid-filled manometer, sometimes referred to as a U-tube manometer, is used to measure the Δp , Equation 13.3 applies.

Liquid volume measurement tanks are calibrated by adding accurately measured volumes of liquid to the tank in increments and measuring the liquid height after each addition. The methods for deriving the relationship of the volume to the liquid height from calibration data are described in Sections 15.1.3 and 15.1.4 of Chapter 15. If the tank is uniform in cross section, the equation will have the linear form

$$h = \beta_0 + \beta_1 V \quad (13.5)$$

where h and V refer to the liquid height and volume, respectively, and β_0 and β_1 are calibration constants. If the form is nonlinear, the best fit to the data may be obtained with a quadratic or cubic equation or a set of two or more linear equations, each representing a segment of the tank.

The equation for calculating volume is $V = (h - \beta_0)/\beta_1$ and, if the differential-pressure method is used to determine the height of the liquid h , the equation becomes

$$V = (\Delta p/g\rho - \beta_0)/\beta_1 \quad (13.6)$$

To apply this equation, the density of the liquid in the tank, ρ (Equations 13.3 and 13.4), and the temperature at the time Δp is measured must also be known.

One of the important steps in a calibration is determining the capacity of the portion of a tank below the zero reading of the liquid level gauge, which is commonly referred to as the "tank heel." After a tank has been put into service in a remote location, it may not be feasible to empty it completely, which is required for a heel measurement.

Failure to recalibrate heels may lead to inventory error because insoluble sediment may collect in the tank. The tank may also change dimensions upon aging, loading, or temperature cycling. A tank must be recalibrated whenever any internal hardware such as heating or cooling coils, suction pipes, and stirrers are replaced. In addition, a periodic recalibration is needed to ensure that the tank walls have not sufficiently distorted to invalidate the calibration equation. Potential sources of bias in volume measurements are:

1. An imprecise or biased calibration, such as if the volume increments added or the liquid heights are not measured accurately, not enough cali-

bration runs (replications) are made, or the temperature of the calibrating fluid was not constant.

2. A change in tank dimensions after calibration.

3. A development of a bias in the liquid level gauge, such as by a movement of the scale on a manometer, a change of density of the manometer fluid, a difference in liquid density between the tank and sight glass contents, and, in the case of electronic instrumentation, drift or instrument malfunction of one of the components.

4. Failure to correct for temperature effects on gauges or the densities of the fluids in the manometer and the tank (Equation 13.3).

13B.4 FLOW MEASUREMENTS

Liquid flow measurements generally have poorer relative precision than volume and mass measurements (flowmeter measurements of volume delivered can be expected to have standard deviations in the 1 to 3% range). Therefore, flowmeters have not often been used for accountability measurements, although they are widely used for process control measurements. With the increasing emphasis on near-real-time accounting and a more prompt detection of losses, however, flow measurements will come into more common use for nuclear material accounting.

Flow measurements fall into two general categories. The first, which also has the greatest precision, is the measurement of discrete volumes per unit time. An example of this type of meter is the common gasoline pump. Other examples are the positive displacement pump (the most precise example of this type of flowmeter), the rotating disk,[†] the rotary vane, and the piston types.

The second category is the velocity or rate type of meter which measures the rate of flow in a continuous stream. Within this general category are subcategories of meters including variable head meters (orifice, venturi tube, pitot tube, flow nozzle, changing area), variable area meters (rotameter, flow drag force, air lift), and velocity and current meters (turbine, thermal, laser-doppler, electromagnetic, and vortex).

Flowmeters are calibrated by passing a known or measured volume of liquid through them at a constant rate in a measured time period. The potential sources of error in calibrations are a pulsating flow or swirling liquid (often caused by pipe bends near the meter inlet), air entrainment, and biased measurements of the calibration volumes. The first two sources of error also apply to measurement applications. In addition, electrical or electronic components and recorders can become biased. Periodic recalibrations are necessary because of the possibilities of wear of moving parts, crud buildup, and leakage of air into the system upstream of the flowmeter.

[†]A sealed disk that "wobbles" with the flow of liquid.

13B.5 CHEMICAL ANALYSIS

Chemical analyses are generally the most precise and accurate of the methods available with which to determine the concentration of an element in nuclear materials. The primary reasons for this are as follows:

1. The material samples can be prepared for analysis by dissolving or pulverizing them to render them homogeneous and separating interfering constituents.

2. The quantity taken for the measurements can usually be near the optimum amount for a precise analysis.

3. The measurements are made in a controlled environment in which the apparatus and instruments can be maintained in an optimum operating condition.

The analytical methods for uranium and plutonium concentrations are the most important ones in nuclear material accounting. In some instances, other element concentrations must be measured to make a correction for interferences with a uranium or plutonium determination.

13B.5.1 Determining Uranium Concentrations

The gravimetric and titration methods are the most widely used analytical procedures for uranium for nuclear material accounting as well as quality control measurements. A gravimetric method is based on measurement by weighing. The usual weighing form of uranium is the oxide U_3O_8 , which is stable in air and does not absorb or lose oxygen at temperatures up to at least 900°C . Furthermore, this oxide is easy to prepare as stoichiometric U_3O_8 , i.e., having the quantitative chemical composition indicated by its formula. Most solutions, salts, oxides, and organic compounds of uranium can be converted directly to U_3O_8 by heating them in air at a temperature of about 900°C . This is referred to as an "ignition." A hydrogen reduction followed by ignition in air is used for maximum accuracy with some compounds of uranium.

The gravimetric method is limited to fairly pure compounds of uranium. Nonvolatile impurities up to about 0.05% by weight are permissible if corrections are made for the weight of the impurity. The concentrations of the impurities are determined by a spectrographic analysis, which is usually performed on the "high grade" uranium compounds in the process for other purposes as well. The principal sources of error in this method are weighing errors; failure to achieve stoichiometric U_3O_8 , which can be caused by using an ignition temperature that is too low or too high; and error in measurement of the impurity concentrations for corrections.

For the direct ignition method, the uranium weight factor ("element factor") is calculated as follows:

$$U_f = \frac{W(F-C)g}{s} \quad (13.7)$$

where

U_f = the uranium factor

W = the number of grams of ignited oxide

F = the nonstoichiometry factor (ratio of stoichiometric U_3O_8 , determined by the hydrogen reduction method, to the uranium oxide obtained by the direct ignition method)

C = the correction for the sum of the weights of impurity oxides, expressed in grams per gram of the ignited oxide

g = the gravimetric factor, $3U/U_3O_8$, calculated from the molecular weights of U and U_3O_8 , corrected for the isotopic composition of the uranium

s = the grams of sample taken for analysis

The F factor is necessary for the highest accuracy. The ignition procedure should also be periodically checked using a standard reference uranium compound, such as one of those available from the U. S. National Bureau of Standards.

There are several titrimetric methods for uranium determinations. They are based on a controlled reduction and oxidation sequence of reactions of uranium ions in an acid solution. The quantity of reagent required to just oxidize the uranium ions is measured, and the quantity of uranium in the sample is calculated from the known reaction stoichiometry. In the case of the coulometric titration method, a controlled electro-reduction is carried out and the "reagent" is coulombs of electricity.

The principal steps for a typical assay of a uranium oxide sample are as follows:

1. Blend the sample. (If it is a granular material, pulverize it first to a fine powder.)
2. Select an analytical subsample to contain between 100 and 300 mg of uranium and weigh it accurately. (Some methods can accept smaller sample sizes.)
3. Dissolve the sample in an appropriate acid, add sulfuric acid, and then evaporate off all acids except sulfuric.
4. Reduce the uranium by adding a measured slight excess of an appropriate reagent.
5. Carry out a catalyzed oxidation of the excess reductant.
6. Titrate the uranium with an oxidizing agent (a chromate solution) of known concentration until the oxidation end point is reached. †
7. Measure the weight or volume of reagent used.

† The endpoint may be detected visually by the color change of a chemical dye that is sensitive to a trace of excess oxidant or electrically by observing an abrupt voltage or current change of an electrode at the endpoint. The electrical indicator method is commonly automated.

8. Calculate the concentration of uranium in the sample as follows:

$$U = \frac{NV(\text{atomic weight of U})/2}{\text{net weight of sample in g}} \quad (13.8)$$

where

U = grams of uranium per gram of sample

N = concentration of dichromate in the titrant in gram equivalents per milliliter of solution (assuming volume measurements used)

V = volume of titrant required to reach the endpoint

The reason for describing the steps of the titration procedure is to show the complexity of a typical analytical titration and the many possible sources of error. They are:

1. Incomplete dissolution of the sample.
2. Weighing errors.
3. Volume measurement errors for the reagents and titrant.
4. Faulty standardization of the titrant.
5. Variation of chemical reaction rates and stoichiometry, as may be caused by improper quantities of reagents or reaction timing and temperatures.
6. Contamination of equipment or reagents.
7. Changes in composition of samples between sampling and analysis.
8. Operator mistakes in detecting the endpoint or making reagent additions.
9. Interferences by other constituents of the sample that react with one or more of the reagents.

The most likely causes of bias are faulty operations, such as incomplete sample dissolutions or timing of reaction steps—use of a working standard that is miscalibrated or has changed composition; change in compositions of samples before analysis; and chemical interferences in the samples. The chemical interferences with which analysts must be concerned vary with the specific titration method used. For the “Davies-Gray” method, which is very widely used, vanadium, manganese, molybdenum, ruthenium, and some of the precious metals are typical of the troublesome elements. When any of these are present in concentrations high enough to cause interference, the analytical procedure is modified to prevent the interference. A description of the various procedural alternatives that may be used to avoid chemical interferences is beyond the scope of this discussion. The reader may consult other manuals, such as that by Rogers (1983).

Very low concentrations and small amounts of uranium are determined by methods more sensitive than titration, such as colorimetric and fluorometric methods. The latter is used most widely, especially for discardable waste solutions.

Uranium is one of the few inorganic species having sufficient natural fluorescence to permit analysis by measuring its fluorescent intensity. Uranium atoms emit a green light when they absorb ultraviolet radiation. The fluorescent intensity varies considerably with the form of the uranium and the matrix.

Two methods of fluorometric analysis for uranium have been developed that achieve high sensitivity and sufficient reproducibility for analysis. They are both used in the nuclear industry for analysis of very dilute solutions such as process effluents. The older method is the fused-pellet method in which uranium salts are fused in a sodium fluoride pellet to enhance the fluorescence (some other pellet materials can also be used). The other method is the pulsed-laser-excitation method. In the latter method the sample is added to a neutral solution containing a pyrophosphate salt as a fluorescence enhancer. Optical instruments with an ultraviolet light source and a photomultiplier tube detector are used in both methods for the quantitative measurements. The calibration of the instrument is accomplished by measuring portions of a solution containing known amounts of uranium and fitting an equation of response voltage versus uranium quantity by the method of least squares.

The principal sources of error in fluorometric analysis are as follows:

1. Instrument drift or instability.
2. Errors in sample volume measurements.
3. Uranium contamination of the fusion pellet, the fusion crucible, or the solutions and containers used.
4. Poor control of fusion temperature or time.
5. The presence of interfering constituents, known as quenching agents in the samples.

Many elements cause fluorescence quenching. The analyst must know when these are present or potentially present and use a suitable variant of the procedure to avoid or correct for the quenching factor. Further information is given by Rogers (1983).

13B.5.2 Determining Plutonium Concentrations

The chemical and coulometric titration methods for plutonium are very similar to those for uranium. The same sequence of operations and the same sources of error are applicable. The primary methodological differences are the oxidation and reduction reagents used, the method of detecting the titration endpoint, and the kinds of materials that cause chemical interferences. When using the amperometric method, which is a widely used procedure, the plutonium results will be high if chromium, manganese, cerium, americium, vanadium, or a few other elements are present. Uranium does not interfere. On the other hand, the results will be low if lead, barium, calcium, selenium, or thallium are present in substantial

quantities. There are method modifications, however, that eliminate most of the interference and in some cases, a chemical separation before analysis is used. [See Rogers (1983) for further discussion and references.]

The coulometric method for plutonium is often used because it requires a much smaller sample size (5 to 15 mg of plutonium is about optimum) than the chemical method. The presence of iron causes high results and a correction is applied that is based on a separate determination of the iron concentration. Some interfering materials that cause low results are zirconium, tantalum, niobium, organic materials (such as those leached from ion exchange resins), fluorides, and phosphates. When any of these are present in any appreciable quantity, the procedure is modified to prevent the interference.

Plutonium analysis bias can also occur if sample solutions are evaporated to excessive dryness, which increases the likelihood of forming plutonium compounds that are very difficult to redissolve, or if the acidity of the samples is too low, which increases the possibility of creating an unreactive polymeric form of plutonium. With the amperometric method, failure to completely oxidize the plutonium or to destroy all excess oxidizing agent before the titration will also cause biased results.

Small quantities of plutonium, such as occur in waste solutions, are generally determined by alpha counting. The isotopic-dilution method, utilizing a mass spectrometer, is also applicable down to moderately low levels, but because of cost, the method is used only when an analysis more precise than α -counting is wanted.

Minute amounts of plutonium can be determined by the alpha-counting method since the specific activity of ^{239}Pu is 1.36×10^{11} alpha disintegrations per minute per gram and alpha counters with background count rates of less than one count/minute and nearly 50% geometry can be used. Semiquantitative determinations of plutonium can be made without calibration of the alpha counters and without exact knowledge of the specific activity of the plutonium (which varies with isotopic composition). Alpha counters are generally calibrated with a standard solution of plutonium, a standard alpha source calibrated with a special absolute alpha counter, or a standardized alpha source of ^{241}Am calibrated by alpha-gamma coincidence counting. The main causes of error in the analysis are self-absorption and scattering losses from counting discs with too much solid material or very uneven deposits of plutonium. Other causes are imprecise calibration of the counter and incorrect isotopic data for the plutonium.

13B.5.3 The Isotopic-Dilution Method of Analysis

In the isotopic-dilution method, a known amount of a tracer isotope of the element to be determined is added to a known amount of sample, the element is separated after steps have been taken to ensure that the tracer and sample constituents are in the same chemical form, and an isotopic

analysis is performed using a surface ionization mass spectrometer. The measured isotopic dilution of the tracer is used to calculate the original concentration of the element in the sample. The method is applied to uranium using ^{233}U tracer and to plutonium using ^{242}Pu tracer. About 1 mg of uranium and 0.01 mg or less of plutonium are suitable for an analysis. The method can be applied simultaneously to uranium and plutonium.

One form of the equation for the isotopic dilution analysis for uranium is

$$C_s = C_t (R_m - R_t) \frac{238}{233} \frac{100}{P} \frac{\text{vol or mass of tracer solution taken}}{\text{vol or mass of sample solution taken}}$$

where

$$\begin{aligned} C_s &= \text{concentration of U in the sample} \\ C_t &= \text{concentration of } ^{233}\text{U} \text{ in the tracer} \\ R_m &= \text{}^{238}\text{U}/\text{}^{233}\text{U} \text{ atom ratio in the mixture} \\ R_t &= \text{}^{238}\text{U}/\text{}^{233}\text{U} \text{ atom ratio in the tracer} \\ P &= \text{wt. isotopic percentage of } ^{238}\text{U} \text{ in the sample uranium} \end{aligned}$$

The R_t value for the stock tracer material and the P value for the sample are determined by isotopic analyses. If the sample is high enriched uranium, the $^{235}\text{U}/\text{}^{233}\text{U}$ ratios rather than $^{238}\text{U}/\text{}^{233}\text{U}$ may be used for the calculations to improve the precision.

The calibration of the method consists of establishing the concentration C_t of the tracer isotope, ^{233}U , in the tracer stock solution and the isotopic ratio, R_t , for the $^{238}\text{U}/\text{}^{233}\text{U}$ ratio of the tracer. The C_t value is determined by analysis of a standard reference solution of uranium whose concentrations of uranium and ^{238}U are known. Such a standard reference solution is prepared by dissolving a weighed quantity of a standard reference material, such as pure U_3O_8 , and making it up to a known volume or mass of solution. The isotopic ratio is determined by mass spectrometry if it was not provided by the calibration laboratory that certified the standard reference U_3O_8 . An accurate tracer calibration is very important. This is done by "analyzing" the standard reference solution in the same manner as unknowns.

The potential causes of error in the isotopic dilution method of determining uranium and plutonium concentrations are:

1. Errors in measuring the volumes or masses of the sample and tracer solutions.
2. Faulty calibration of the tracer solution.
3. Changing concentration of the tracer solution, such as by evaporation.
4. Failure to achieve complete chemical and isotope homogeneity of the mixture.
5. Inaccurate isotopic ratio measurements.

There are several calibration requirements associated with the isotopic analysis instrumentation. Section 13B.8 contains a summary of these.

13B.6 SAMPLING

Analytical results are of little value when the composition of the samples analyzed are not representative of the material from which the sample was withdrawn. The representativeness of a sample depends on both the homogeneity of the material being sampled and the sampling method. Samples of homogeneous material tend to be representative regardless of what portion of the material is taken as the sample. When the homogeneity of a material has not been firmly established, however, the method of sampling must be planned with care. The general principle is that to be representative, every portion of the bulk material must have an equal chance of being included in the sample.† That is not always easy to achieve in practice. When it cannot be made part of the sampling plan, the material must be blended until homogeneous before sampling.

Bulk nuclear materials are sampled in a variety of ways. The following methods are typical:

1. Grab or dip sampling—dipping or scooping a portion of the material from the top of the container. A petcock method of withdrawing a sample of liquid from a tank is equivalent to grab sampling.
2. Thief sampling—a closed bottle or tube is lowered into the liquid or powder to the desired level and then opened for filling.
3. Core sampling—an open-ended tube is inserted into the liquid or powder to the desired depth, then withdrawn with the sample that fills the tube. For liquids and free-flowing solids, a check valve is attached to the lower end to close the tube and retain the sample while the sampler is withdrawn.
4. Flowing-stream sampling—a stream splitter, outlet, or petcock withdraws a small portion of a moving material from a pipe, conveyer, chute, or spout. The withdrawal is continuous or at regular intervals during a given time for a sample representing the material transferred during that time. Automatic mechanical devices of many kinds are used. Some withdraw a quantity that is proportional to the flow rate of the material and are called **proportional samplers**. One type of flowing stream sampler, referred to as a **spinning disc sampler**, is a rotating device that captures a small increment (a complete cross-sectional portion) of a flowing stream at frequent intervals.
5. Riffle sampling—a quantity of a free-flowing material is subdivided by a splitter into smaller portions, one of which is taken as the sample for analysis. The method is commonly used in laboratories to select a representative analytical sample from a larger sample of the plant material.

†This does not guarantee a small sampling random error but simply ensures that the expected sample composition is the mean of the bulk material composition.

Grab or dip sampling is the simplest and most inexpensive method, but it should not be used for nonhomogeneous materials. The flowing-stream and riffle-sampling methods are least subject to sampling errors with such materials. The other methods must be used with discretion when the material is not homogeneous because if a random selection of the sampling position for thief and core samplers is not made, the samples may be nonrepresentative of the material. Even a core sample from a can of powder can be nonrepresentative if the probe does not reach to the bottom or if some part of the contents (e.g., the outer circumference) is never probed.

The most effective way to avoid both random error and bias in sampling is to blend the material to a homogeneous state before sampling. Liquids are easily blended by stirring, air bubbling, or convection (by differential heating). Some process liquids, however, are not blended when sampled because of cost, time constraints, or equipment limitations. Other common constraints are a lack of stirring equipment in storage vessels—especially those with constrained geometries for nuclear criticality prevention—and mixing difficulties with solutions in evaporators, extraction columns, and tanks filled with neutron-absorbing material such as borated raschig rings.

Powders, fuel pellets, metals, and alloys may have significant composition differences between containers, between batches, or between lots. In nuclear fuel fabrication, uranium and plutonium oxide powders are blended before pellet formation to create homogeneity within lots. Cans or trays of pellets from the lot are also cross-blended before fuel rod loading. These kinds of materials, however, might be sampled for inventory accounting purposes at intermediate stages of the process when they have not been blended. In those cases, the sampling plan must make provisions for the nonhomogeneity problem. The most common solution is to withdraw many subsamples from the batch, such as one from each container, and composite the subsamples for the analyses.

Analysts must also be aware of potential sources of sampling error that can take effect after the primary sample has been taken from the bulk material. There are two classes of such errors:

1. Inadequate protection of the sample to prevent composition changes while awaiting analysis.
2. Failure to attain homogeneity of the sample before withdrawing the portion to be used for an analysis.

Protection of the sample, which is usually a containment issue, is the joint responsibility of the person who prepared the sampling plan and the person who withdraws and packages the sample. The analyst, however, should be alert to containment problems and should notify the recipients of the analytical results of any observed deficiencies. Analysts should know if incoming samples are composites of subsamples that could differ in compo-

sition and blend them accordingly. In addition, if the sample contains granular or agglomerated material, it may be necessary to pulverize the material before blending it. If blending is not complete, the portion of the sample taken for analysis may not be representative of the whole. Analyses of two separate portions will reveal such a problem.

13B.7 NONDESTRUCTIVE ASSAY

Nondestructive Assay (NDA) is the measurement of a characteristic of an item that can be performed without altering the item. Most NDA methods used in the nuclear industry are based on comparison of the radiation or the heat generation from an unknown quantity of the element of interest to the radiation from a known quantity of the element. Both **passive and active measurement techniques** are used, where passive NDA means that the characteristic radiation spontaneously emitted by the isotope is measured, whereas active NDA means that the radiation emitted is induced by an external source, generally a neutron source. Most of these two classes of methods are based on quantitative measurement of gamma ray, X-ray, or neutron radiation, including coincident neutron and coincident gamma-neutron radiations. Another type of NDA method is calorimetry, in which the heat production rate from radioactive decay of a nuclide is measured and related to the quantity of the nuclide in the item.

One of the key features of these NDA methods is that they are used to measure the total quantity of specific isotopes of uranium or plutonium in an item or container of material rather than concentration. (Measurements can be performed on representative samples of the material and concentration values for the element of interest obtained, but these are not discussed here.)

An NDA method instead of a destructive analysis method is most commonly used when one of the following circumstances exists:

1. A more rapid or economical assay than can be achieved by a laboratory analysis is needed. This is common for process control measurements.
2. Destructive analysis or sampling is either not feasible or is too costly. Assay of fuel rods is an example.
3. A representative sample is not obtainable because the material is very heterogeneous. Typical examples are production and laboratory wastes and certain scrap materials.

Although chemical analysis methods are capable of greater accuracy and smaller standard deviations for uranium and plutonium concentrations than NDA methods, some NDA methods are close competitors when the material form and composition are known and are very consistent. In most applications, the standard deviations of NDA methods are much larger than those typical of the chemical methods. It must be recognized, how-

ever, that when material heterogeneity is a cause of poor NDA accuracy, a conventional chemical analysis may not be better unless the material is first blended sufficiently to permit representative sampling. This would often require pulverization or dissolution of the entire quantity to prepare it for sampling. In many of the applications of NDA, the larger uncertainties can be accepted because either the category of material involved is not a major component of the material balance or the assay results will eventually be backed up by other analyses.

NDA data will often be very precise; that is, the measurement-to-measurement data for a single item will be extremely reproducible. The main contributions to reproducibility are instrumental stability and the reproducibility of the counting geometry. The assay result, however, may be biased, or the item-to-item error in assays of different items of the same category (i.e., all subject to the same set of calibration data) may be large. There are several important causes of this kind of error:

1. The presence of materials in the items that strongly absorb (attenuate) the radiation to be measured or the stimulating radiation (e.g., the source neutrons for an active NDA method) so that large attenuation corrections are required.
2. Variable quantities of radiation absorbers from item to item.
3. The presence of dense forms (agglomerates or large particles) of the nuclide being measured, causing self-attenuation.
4. Nonuniform distribution of the nuclide and matrix materials, particularly for NDA instruments that do not compensate by integrating the count rates over many positions of the target, e.g., by rotating the item while counting.
5. A mismatch between the average composition of the items and that of the calibration standards.

The NDA methods for uranium and plutonium determine the quantities of specific isotopes, not the element quantity itself. For example, passive gamma assay of uranium is based on the 185 keV emissions of ^{235}U , and the passive neutron count assay of plutonium is based on the combined neutron emissions by spontaneous fission of ^{240}Pu and the α, n reactions with the light elements present, where the primary alpha source is ^{239}Pu . Therefore, the isotopic composition of the element must be known to calculate the element quantity from an NDA result, or the calibration standards must be made up of quantities of the element having an identical isotopic composition.

Calibration of NDA methods requires material standards that are very similar to the unknowns to be measured. This similarity extends to the composition, distribution, material form and container size, shape, material, and wall thickness. Calibration standards are prepared by mixing known quantities of uranium or plutonium with matrix material like that

of the unknowns to be measured. These are packaged in the same type of container as used for the unknowns. Furthermore, the uranium or plutonium should have the same chemical form and particle size as that expected in unknowns. Standards are constructed for each type of unknown expected, and the range of uranium or plutonium contents should span the full range of intended use.

13B.8 ISOTOPIC ANALYSIS

The usual method for determining the isotopic compositions of uranium and the transuranium elements is mass spectrometry. A mass spectrometer utilizes electrostatic and magnetic field interactions with charged particles to separate the particles according to mass and to measure the quantity at each mass. The sample size required for an analysis is 10^{-8} to 10^{-5} g of uranium or plutonium, depending on the sensitivity of the instrument. Two types of mass spectrometers, the gas-source instrument and the surface-ionization instrument, are used routinely for uranium isotopic analysis. The gas instrument is used for UF_6 samples and is currently utilized primarily in uranium isotope enrichment plants where large numbers of such analyses are required. In the remainder of the nuclear industry, as well as in most nuclear research establishments, the more versatile surface-ionization instrument is used.

Samples for isotopic analysis by surface ionization are dissolved in an acid, and the element (uranium or plutonium) is purified to eliminate other elements that may interfere with a precise analysis. A dilute, slightly acid solution (nitric acid) of the element is prepared, and a fraction of a drop of the solution is dried and ignited on a tungsten or tantalum filament. The filament is placed in the source of the mass spectrometer and heated to a temperature that causes a steady emission of positive ions of the element. All of the isotopes of the element appear in the ion beam in the same ratios as they occur in the sample.† The mass spectrometer system is calibrated by running isotopic standards that are obtainable from the U. S. National Bureau of Standards. Direct ion ratio determinations by surface ionization mass spectrometry are very nearly correct without the use of calibration standards, but such a procedure is not considered sufficiently accurate for nuclear material accounting.

Four potential sources of isotopic analysis bias are:

1. Miscalibrated or nonlinear electronic circuitry in the ion collector, amplifier, and readout systems.
2. Isotopic discrimination (isotopic fractionation) occurs in ion production and in focusing the isotope beams, and electron multiplier detectors have slightly different response sensitivities for different isotopes.

† A small isotopic fractionation occurs in the ion source, but a correction factor derived by calibration is applied.

3. Interferences by extraneous ions having the same mass values as the element isotopes and impurities that cause erratic ion production rates or high vapor pressures in the source region, resulting in beam scatter and poor resolution.

4. Contamination of filaments, laboratory glassware, or reagents by the same element but with a different isotopic composition.

Isotopic analysis bias is avoided by monitoring the instrument performance frequently with a reference isotopic standard. In addition, certain components of the instrument are calibrated and tested periodically. Tests are made to demonstrate that the sensitivity ranges of the detector amplifier are accurate and linear, the electron multiplier gain is constant, and the discrimination correction factor of the detector system has not changed. Most of these tests and calibrations are made with isotopic standards.

REFERENCES

- Brouns, R. J., J. A. Merrill, and F. P. Roberts, 1980, *Methods of Determining and Controlling Bias in Nuclear Material Accounting Measurements*, NUREG/CR-1284 (PNL-3264), Nuclear Regulatory Commission, Washington, D. C. 20555.
- IAEA, 1980, *IAEA Safeguards: Guidelines for States' Systems of Accounting for and Control of Nuclear Materials*, IAEA/SG/INF/2, International Atomic Energy Agency, Vienna, Austria.
- Korstad, P. A., 1980, *Accounting Systems for Special Nuclear Material Control*, NUREG/CR-1283 (PNL-3265), U. S. Nuclear Regulatory Commission, Washington, D. C. 20555.
- Rogers, D. R. (Ed.), 1983, *Handbook of Nuclear Safeguards Measurement Methods*, NUREG/CR-2078, U. S. Nuclear Regulatory Commission, Washington, D. C. 20555.

CHAPTER 14

Modeling Measurement Variability

14.0 INTRODUCTION

A brief description of the modeling process for describing the effect of measurement variability on a material balance, or one or more components of the material balance equation, was given in Section 13.3.1. In this chapter, a more complete description is presented of the basic statistical concepts involved in modeling and estimating the effect of the presence of measurement error on accounting results.

14.1 MODELING MEASUREMENT SYSTEMS

Working down through the accounting and measurement processes is necessary before the **experiments** (in this case, mostly measurement procedures) that generate the basic **random variables** can be specified. In Section 2.2, a random variable is defined as “a numerically valued function defined over a sample space.” In simpler terms, it is a rule which assigns a numerical value to each possible outcome of an experiment. For our purposes, the experiment is the application of a measurement system to an item containing nuclear material, and the numerical value is the measured content in appropriate units. The concept of a random variable is a useful and convenient way of describing the nondeterministic, or stochastic, nature of the measurement process. The important idea is that measurements are not perfectly repeatable. This chance behavior of random variables can be described in part by specifying their expected values and variances, or completely by specifying a distribution function. Repeated runs of the experiment produce observed sample values of the random variable from which the behavior of the random variable can be estimated.

14.1.1 The Basic Approach

Table 14.1 is a collection of definitions pertinent to modeling measurement systems. The definitions of the American National Standards Institute (ANSI N15.5-1982, *Statistical Terminology and Notation for Nuclear Materials Management*) are given along with references to the appropriate section of this book and the ANSI N15.5 standard. Usually the two definitions carry essentially the same meaning. They are provided here because they reflect the understanding of these terms in the nuclear industry.

TABLE 14.1
ANSI Definitions Relative to Modeling
Measurement Systems

| Definition | Text
section | ANSI
N15.5
section |
|--|-----------------|--------------------------|
| 1. Population: A collection of objects or events. | 3.0 | 2.3.1.1 |
| 2. Random variable: A variable that takes on any one of the values in its range according to a specific, but usually unknown, probability distribution. | 2.2 | 2.3.1.5 |
| 3. Parameter: An unknown constant that is associated with, or is used to characterize, a distribution or density function. | 3.0 | 2.3.1.10 |
| 4. Estimator: A function of a sample (X_1, X_2, \dots, X_n) used to estimate a population parameter. | 3.0 | 2.3.6.1 |
| 5. Expected value: The value obtained by multiplying the value of a random variable by the value of the density function of the random variable and summing or integrating the product over the range of the random variable. | 2.3 | 2.3.1.12 |
| 6. Sample: (a) A collection of elements selected from a population according to some sampling plan. (b) To select or choose a collection of elements from a population. | 8.0 | 2.3.3.2 |
| 7. Estimate: (a) A particular value, or values, realized by applying an estimator to a particular realization of a sample, i.e., to a particular set of sample values (x_1, x_2, \dots, x_n). (b) To use an estimator. | 3.0 | 2.3.6.2 |
| 8. Uncertainty: A concept used to describe the inability of a measurement process to measure exactly the true value sought. | - | 2.3.13.12 |
| 9. Error: A deviation from the true value sought. | 14.1.3 | 2.3.13.3 |
| 10. Random error: (a) The deviation of a random variable from its expected value. (b) A deviation which occurs "randomly" (according to a probability distribution) over replicate measurements. | 14.2.1 | 2.3.13.7 |
| 11. Bias: (a) The deviation of the expected value of a random variable from a corresponding correct value. (b) An error which remains constant over replicate measurements (synonyms: deterministic error, fixed error, systematic error). | 14.2.1 | 2.3.13.8 |

(Table 14.1 continued on next page).

TABLE 14.1 (Cont'd)

| Definition | Text
section | ANSI
N15.5
section |
|---|-----------------|--------------------------|
| 12. Variance: (a) (population) The expected value of the square of the difference between a random variable and its expected value. (b) (sample) The sum of the squared deviations from the sample mean divided by one less than the number of values involved. | 2.3 | 2.3.5.1 |

The basis for the application of statistics to material accounting rests on this approach to modeling measurement systems. Quantitative conclusions about the real world of material accounting depend on the statistical **populations** (Definition 1) resulting from the application of the measurement systems to process items or measurement control standards. Each such application generates a **random variable** (Definition 2). Each random variable can be characterized by a distribution function with certain **population parameters** (Definition 3). Depending on the reasons for collecting data, **estimators** (Definition 4) are selected to help answer the questions of interest. The **expected values** (Definition 5) of the estimators can be determined from general statistical theory. The quantities of major interest in safeguards are the means and variances of the random variables. Much of this modeling process can be done without looking at any measurement data. The distributional assumptions involved can be checked after the data have been collected, or past experience can be used to provide reasonable starting assumptions.

The actual work of determining numerical values for the parameters that describe the populations, or their means and variances, is done by collecting **samples** (Definition 6) of observations from the populations of possible measurements on process items and standards. These outcomes of the measurement process serve as the only connection with the true quantities present. For process items, only a sample of size one may be available from the population of possible measurements, like the weight of the UO_2 in a can of UO_2 product. On the other hand, 15 or 20 measurements of a measurement control standard weight may be taken during an accounting period. Such measurements taken over time make it possible to determine whether or not the weighing system is generating random variables from the same distribution—i.e., whether or not each observation comes from a distribution with the same expected value and variance which is characteristic of the weighing system. Sample estimates from such data can be compared with standard values, or possibly historical values, of the parameters of the distribution to determine whether or not the weighing system is out

of statistical control. Procedures for estimation and control of measurement parameters are the subject of Chapter 15.

The process of working up from the variation in the basic measured values to the variability in the estimated inventory difference recognizes that the accounting process is an estimator of an observed value of the inventory difference, which is a function of the basic measurement random variables and is, therefore, itself a random variable. Similarly, the estimated variance of the estimated inventory difference is also an observed value of a random variable derived from other random variables through algebraic relationships determined by the operational definition of the inventory difference estimator. Although the statistical distributions associated with the basic measurement errors are frequently well characterized, analytical characterization of the distribution of the inventory difference may be difficult without numerous simplifying assumptions. Definitions 9 through 12 of Table 14.1 characterize specific errors of measurement in the measurement models and their behavior.

In summary, the basic operational steps involved in modeling and estimating the effect of measurement variability on a material balance or a material balance component are:

1. Observing and analyzing the measurement operations involved in the accounting process, and modeling both individual determinations of material quantity and their combination into accounting results in terms of the measurement error effects.
2. Deriving from this model algebraic expressions for the variability in the accounting results in terms of the parameters (usually expected values and variances) describing individual measurement errors.
3. Specifying the manner in which the parameters characterizing the measurement errors can be obtained from measurements on process items and control standards, or by other experiments specifically designed to isolate and estimate individual sources of measurement error.

The estimation procedures used in step 3 to characterize measurement error are not necessarily unique to material accounting. Under appropriate conditions the results can be extended to other uses of the measurement procedures involved, instead of being restricted to a specific material balance.

This chapter deals with the definition of the error effects associated with measurement systems, their propagation through the accounting process, and the derivation of appropriate estimators of the uncertainty in material accounting results. If an exact model of the specific measurement and accounting system is required, this process can become quite complex. Because of this, a somewhat simplified generic model of the effect of measurement uncertainty on accounting results was developed that provides an adequate approximation for most evaluation purposes. This model

is the basis of the procedures described in the rest of this chapter. Appendix 14A contains a specific example of a more complete modeling process. The need for more precise estimates in specific applications may warrant such exact treatment. The adequacy of the generic treatment can only be validated by periodic comparison with a more exact treatment.

14.1.2 The Accounting Process

As noted in Section 13.3.1, an observed inventory difference is usually a complex combination of measurement results on a wide variety of process materials involved in a particular production process. These results and the measurement uncertainty they contain must be described so as to expedite the correct and complete determination of estimates of the total variability in the basic material balance components. Two different descriptions are needed, a description of the accounting process and a description of the error structure in the measurement results. Rarely are both descriptions needed at any given step in the modeling process, but it is necessary to recognize their existence and keep the distinction clear. Even with careful definition of the notation used, it is frequently difficult to maintain clarity with respect to the particular structure involved in a given description or analysis.

The three basic measurement results that may be involved in determining the quantity present in a given unit of nuclear material are the bulk amount (weight or volume) of material, the concentration of the element being accounted for and, when required, the concentration of a particular isotope of the element. These results will be designated by y_1 , y_2 , and y_3 , respectively. The accounting process is carried out on the amounts $x_1 = y_2 y_1$ or $x_2 = y_3 y_2 y_1$ of element and isotope, respectively. A given inventory difference will involve only one of these quantities, and the breakdown into measurement results is necessary only to determine the measurement uncertainty, not to describe the accounting process.

Occasionally an additional subscript is needed to show that observed results are differences between individual measurements—i.e., a net weight as a difference between a gross weight and a tare weight, a net count as a difference between a total count and a background count, or a net transfer as a difference in the content of a tank before and after the transfer. Conversely, there will be instances when the element or isotope amount is measured directly by nondestructive assay (NDA) rather than determined from separate measurements of bulk amount and concentration. When an isotope amount is measured directly, as may be the case with NDA, the element amount is determined by dividing the measured amount of isotope by the isotope concentration factor.

To characterize the accounting structure and associated measurement error effects, it is most convenient to use the notation for the analysis of

variance with nested hierarchical models (Section 5.3). Because by definition a stratum is a collection of similar material for which the observed results are obtained using a fixed set of measurement procedures, it is the natural unit whose uncertainty can be determined in terms of a fixed set of measurement error variances. Any of the amounts or concentrations in the previous paragraph may be determined on the k^{th} item in the j^{th} batch in the i^{th} stratum. When it is necessary to distinguish between a subscript indicating the type of measurement and the subscripts designating a particular item, batch, or stratum, a comma will be used to separate the subscripts. Thus, $y_{1,231}$ designates the bulk weight of the first item of the third batch of Stratum 2; and $x_{1,12}$ designates the amount of element in the second batch of Stratum 1.

As with the estimation processes for an analysis of variance with nested hierarchical models, the number of items and total amounts of material in the batches and strata play an important part in the analysis. In this chapter the notation used in Section 5.3 and defined in Section 5.1.1 is modified slightly by eliminating the dots used to indicate the absence of the subscript in the sum or average over a particular subscript. For nested hierarchical models, there should be no confusion because the order of summation is always the reverse of the hierarchical order. Thus, if $x_{i,j,k}$ is the amount of element in the k^{th} item in the j^{th} batch of the i^{th} stratum, then $x_{i,j}$ is the total amount in all items in the j^{th} batch of the i^{th} stratum and x_{i} is the total content of all batches in the i^{th} stratum. Again note that in considering a single inventory difference involving either the amount of element or the amount of isotope, but not both, the initial subscript and the comma could be dropped because no confusion would result. N_{ij} and N_i designate the number of items in the j^{th} batch of the i^{th} stratum and in the i^{th} stratum, respectively; and a_i will designate the number of batches in the i^{th} stratum, so that

$$N_i = \sum_{j=1}^{a_i} N_{ij} \quad (14.1)$$

14.1.3 The Error Structure

The result of a measurement process will always differ to some degree from the true value of the characteristic being measured. This difference, usually referred to as a measurement error, arises from a combination of individual error effects associated with the various components of the measurement process. These sources of error, or sources of variability, must usually be considered individually to determine the cumulative variability in a measurement result.

It is usually impossible to specifically identify and account for all factors which might affect a measurement process. This does not imply that the contributions of some factors should be ignored, but rather that they

be combined with others for consideration in the statistical analysis. A natural way to classify factors is according to the component of the measurement process with which they are associated. Some measurement process components which are frequently used to characterize individual sources of variability are the following:

1. **Material sampling:** Samples of material are physically removed from a bulk amount to measure the element or isotope concentration (or some other characteristic of interest). The value of the concentration obtained depends upon the particular sample of material chosen.

2. **Statistical sampling:** A subset of items to be measured is selected from the population of items for which the total amount is required (Section 8.1). The measurement result for the characteristic of interest which will be applied to the whole population is affected by the sample of items selected.

3. **Analytic technique:** The measurement obtained from a measurement process may depend upon the analytic technique or procedure used to measure the material.

4. **Instrument:** The instrumentation used to obtain a measurement may affect the result. Variations in results may occur due to variations in line voltage, temperature, etc.

5. **Analyst or Operator:** The people operating a measurement process may affect resulting measurements. All four of the sources discussed above might involve human operators or analysts. Depending upon the magnitude of the human effects, it may or may not be preferable to separate them from their process component counterparts.

6. **Calibration:** Using a calibration equation or procedure to indirectly measure a quantity of interest in a measurement process affects the measurement obtained.

7. **Environmental conditions:** The environment surrounding a measurement process may affect measurements.

Each of these sources for an individual measurement can be associated with each measurement involved in determining the accounting result for a particular item.

As noted previously, generic treatments of the error structure do not specifically identify all of the individual sources in every model of the error structure used to determine the variability in the accounting process. Jaech (1973) uses a breakdown based on five general measurement operations normally associated with each individual item amount: bulk weight or volume determination, sampling for element, analysis for element, sampling for isotope, and analysis for isotope. Later treatments (Argentisi, Casilli, and Franklin, 1979; IAEA, 1982; Messinger, Lumb, and Tingey, 1981) were based on only three sources of variability—bulk measurement method, sampling procedure, and analytical procedure—because, as noted

earlier, any given calculation of variability in accounting results involves only element amount or isotope amount. Also, the determination of element or isotopic concentration can frequently be considered as a single source of variability combining sampling and analysis.

14.2 MEASUREMENT ERRORS

In this section, a model of the error effects associated with a specific item amount is formulated. The combination of these effects to determine the variability in accounting totals will be considered later.

14.2.1 Random Errors and Biases

The simplest possible model for a single measurement on a single item is

$$y = \mu + \Delta \quad (14.2)$$

where μ is an unknown parameter representing the true value of the measurement and Δ is the error of measurement. If Δ is assumed to be a random variable whose expectation is β , then the observed result can be written as

$$y = \mu + \beta + \epsilon \quad (14.3)$$

where ϵ is a random variable whose expectation is zero and β is an unknown constant deviation of the measurement from the true value. For the specific stipulation of the nature and extent of the measurement process or procedures producing the set of y values with respect to which the expectation is defined, ϵ is the random effect (Definition 10, Table 14.1) and β the bias, or systematic error (Definition 11). The difficulty in distinguishing between these two types of contributions to the measurement error arises from the difficulty in determining those parts of the measurement operation, time periods or external conditions which together define a specific random variable with constant expectation and variance or with a specified distribution. The error in a weighing operation may have a different expectation from day to day depending on temperature. In that case, the error distribution for a week's results is not the same as for a given day's results. The variance in the errors associated with an analytical procedure may differ depending on the operator carrying out the procedure, so that the error distribution for the results obtained by all operators will differ from that for the single operator assigned to the operation on a given day. The problems of interpretation caused by this complexity are discussed more completely in Section 14.3.1.

Generally, the more inclusive the treatment of the sources of variability in a measurement process becomes, the smaller will be the residual fixed effect and the larger the fraction of the error assignable to the various sources of random effects included in the difference Δ . The traditional underestimation of measurement error from repeated measurements is generally due to the lack of replication in the measurement process of all sources of error in the measurements that contribute to the total variability. Conversely, the errors in the average or sum of repeated measurements, for which not all sources of variability are replicated, will in general be larger than if all measurement operations were repeated for each measurement.

Suppose that a specific measurement process or operation represented by the simple linear model of Equation 14.3 is repeated on a series of items with true values μ_k , $k = 1, \dots, N$. Then

$$y_k = \mu_k + \beta + \epsilon_k \quad (14.4)$$

where by definition, $E(\epsilon_k) = 0$ for all k . When the specific distribution of the error effects is not known, it has been customary to consider either the effects ϵ_k or the relative effects $\eta_k = \epsilon_k/(\mu_k + \beta)$ arising from the replication of an identical measurement operation during a specific time period as identically distributed random variables. When the replicated errors are relative, substituting $\epsilon_k = \eta_k(\mu_k + \beta)$ in Equation 14.4 gives

$$y_k = (\mu_k + \beta)(1 + \eta_k) \quad (14.5)$$

If $\text{Var}(\epsilon_k) = \sigma^2$ and $\text{Var}(\eta_k) = \delta^2$ for all k , Equations 14.4 and 14.5 are the models for replicated measurements with constant **absolute** variance and constant **relative** variance, respectively. The variability in the error has been described in terms of the expectation β and either the variance σ^2 or the variance δ^2 . When a specific distribution is known or can be reasonably assumed, an exact treatment is possible, although in some cases it may be complex. For example, if all or part of the error is the result of counting random emissions, as is frequently the case in NDA measurements, that part of the error due to the random variation in the number of emissions in any given time period will have a Poisson distribution. In this case $\text{Var}(\epsilon_k) = \mu_k$, and the errors have neither constant relative nor constant absolute variance.

The parameter β could also have been assumed to be a constant multiple $\beta'\mu_k$ of the true value rather than an absolute quantity. In this case Equation 14.5 becomes

$$y_k = \mu_k(1 + \beta')(1 + \eta_k) \quad (14.6)$$

Taking the natural logarithms of both sides,

$$\ln y_k = \ln \mu_k + \ln(1 + \beta') + \ln(1 + \eta_k) \quad (14.7)$$

For β' and η_k small relative to 1,

$$\ln y_k \cong \ln \mu_k + \beta' + \eta_k \quad (14.8)$$

Thus, constant relative effects in the original measurements are approximately equivalent to constant absolute effects in the logarithms of the measurement results. This relationship is particularly convenient when estimating relative error variances from repeated measurements because the methods of Chapters 5 and 6, based on linear additive models, can be applied directly to the logarithms of the measurement data.

A more general model of replicated measurements allows for identically replicated error effects that contain both absolute and relative components. The appropriate model is

$$y_k = (\mu_k + \beta + \beta'\mu_k)(1 + \eta_k) + \epsilon_k \quad (14.9)$$

This model is probably the most realistic portrayal of the actual situation with respect to the error in safeguards measurements. The possible presence of both absolute and relative random error effects increases the complexity of procedures for determining the expected variability in accounting results. Most of the existing treatments assume relative measurement errors associated with each of the various sources of variability, but extensions to accommodate both relative and absolute components are not difficult. When choosing models for estimating error variance components from available control data, the choice should be made on the basis of which assumption leads to a more nearly constant variance for each of the classes of random effects in the additive model. Some of the combinations which frequently occur are:

1. An additive combination of absolute effects.
2. An additive combination of relative effects.
3. A multiplicative combination of relative effects.

Sometimes it may be necessary to use the general model of Equation 14.9 and perform a non-linear analysis if none of these assumptions is appropriate. Regardless of the model chosen to represent the available data from

which the variance components are to be estimated, it is the individual relative or absolute variance components resulting from the analysis which are used to compute the variability in the accounting results.

The assumption of relative error variances as the basis for calculating the total variability is largely a matter of convenience, although in general the errors in chemical analytical measurements tend to be relative rather than absolute. The actual accounting results are determined as products of bulk quantities, element concentration, and isotopic concentration. It was noted previously that relative error effects have an approximately additive effect on the logarithm of the results. Thus, relative error variances will propagate approximately additively if the equations of Section 2.7 are used to deal with products. As will be seen in the next section, this simplifies the process for estimating the total uncertainty in an accounting quantity due to the several measurements involved in a single accounting result. The validity of this assumption can be enhanced by associating the relative error variances with strata within which differences in the item amounts are small. In this case, there is little difference in whether absolute or relative errors are assumed. When $\mu_k = \mu$ for all k , Equation 14.9 reduces to

$$y_k = \mu' + \epsilon'_k \quad (14.10)$$

where $\mu' = \beta + \mu(1 + \beta')$ and $\epsilon'_k = \mu'n_k + \epsilon_k$. It follows that under this condition, the variance components associated with relative and absolute sources of bias effects and random error effects cannot be estimated separately, and are indistinguishable when estimating total variability. Finally, when the variations in the true measurements are large, most measurement operations are designed to maintain the relative, rather than the absolute, error of measurement in the varying bulk quantities or concentrations.

Throughout the entire process of modeling the contribution of error sources to the total variability and estimating individual components, the compromise is between realism in the modeling of individual sources of measurement error and the difficulty of deriving exact expressions for the total variability arising from the complex variety of error sources which affect an accounting result.

14.2.2 Errors in Item Amounts

The amount of element or isotope present in an accounting unit is usually the product of two or three measurement results, each of which may be affected by more than one source of measurement error. The problem of multiple contributions to a single result is considered first, and then the combination of measurement errors into the total error in an item amount is considered.

For a given measurement result the errors β , β' , ϵ , or η may be a combination of effects arising from various parts of the measurement operation. For example, the element concentration y_2 may contain errors due to sampling, aliquoting, extraction, titration, or chromatography. In this case

$$\epsilon_2 = \sum_q \epsilon_{2q} \quad (14.11)$$

where the ϵ_{2q} are random variables with zero expectation contributing to the total error in an individual measurement result. If these contributions are assumed to be independently distributed, then it follows from Equation 2.123 that

$$\sigma_2^2 = \text{Var}(\epsilon_2) = \sum_q \text{Var}(\epsilon_{2q}) = \sum_q \sigma_{2q}^2 \quad (14.12)$$

A similar result holds approximately for independent sources of relative error in a single measurement if the errors are small. If

$$1 + \eta_2 = \prod_q (1 + \eta_{2q}) \quad (14.13)$$

then neglecting products of two or more η 's leads to the result

$$\eta_2 \cong \sum_q \eta_{2q} \quad (14.14)$$

Again, from Equation 2.123 it follows that

$$\delta_2^2 = \text{Var}(\eta_2) \cong \sum_q \text{Var}(\eta_{2q}) = \sum_q \delta_{2q}^2 \quad (14.15)$$

The bias effects β and β' combine additively, so that

$$\beta_2 = \sum_q \beta_{2q} \quad (14.16)$$

is the total bias effect from the combined contributions of absolute bias effects, and

$$\beta'_2 \cong \sum_q \beta'_{2q} \quad (14.17)$$

is the approximate combined contribution of the relative bias effects when they are small relative to unity.

In determining the errors in the amount of material in an accounting unit the two quantities of interest are usually

$$x_1 = y_1 y_2$$

$$x_2 = y_1 y_2 y_3$$

For the case of relative effects, applying Equation 14.6 to each y_p gives for either x_1 or x_2

$$\Pi_p y_p = \Pi_p \mu_p (1 + \beta'_p)(1 + \eta_p)$$

For small ϵ_p and β'_p , expanding and dropping terms containing products of two or more errors leads to the approximate result

$$\Pi_p y_p \cong \Pi_p \mu_p (1 + \sum_p \beta'_p)(1 + \sum_p \eta_p) \quad (14.18)$$

This result also can be obtained by noting that

$$\ln(\Pi_p y_p) = \sum_p \ln y_p$$

so that applying Equation 14.8 to each $\ln y_p$,

$$\ln(\Pi_p y_p) \cong \sum_p \ln \mu_p + \sum_p \beta'_p + \sum_p \eta_p$$

Exponentiating, and remembering that $e^x \cong 1 + x$ for x small, produces the result given in Equation 14.18.

Because $\mu = \Pi_p \mu_p$ is the product of the true values of the measurement results, and hence the true amount of material present in an item, Equations 14.14, 14.17, and 14.18 can be combined to obtain

$$\begin{aligned} x &\cong \mu(1 + \beta' + \eta) \\ &= \mu + \mu\beta' + \mu\eta \end{aligned} \quad (14.19)$$

where x is the amount of element or isotope in a single item, and

$$\beta' \equiv \sum_p \sum_q \beta'_{pq} \quad (14.20)$$

$$\eta \equiv \sum_p \sum_q \eta_{pq} \quad (14.21)$$

This result greatly simplifies the statistical treatment of the accounting combination of unit quantities into accounting totals to be considered in the next section. In the first place, the total error can be considered as the sum of the corresponding individual error effects for all error sources affecting a particular estimated amount without regard to the particular measurement involved. In the second place, the only distinction between the treatment of element amounts and isotopic amounts becomes the sources of variability present; hence, the general treatment of error combination need not distinguish between the amounts x_1 and x_2 with respect to the formulas and techniques required.

In summary, the relative error in the amount of element or isotope in each item is approximately an additive combination of relative measurement errors arising from the particular set of sources of measurement variability $r = 1, 2, \dots, m$ affecting that amount. Equations 14.20 and 14.21 can be rewritten in the form

$$\beta' \equiv \sum_r \beta'_r \quad (14.22)$$

and

$$\eta \equiv \sum_r \eta_r \quad (14.23)$$

14.2.3 Errors in Stratum Totals

In this section the consideration of error effects is extended to a single stratum containing $N_i = N$ items. If the stratum consisted of a single batch of N items (or of N batches with $N_j = 1$ for all batches) whose content had been determined by measurement operations replicated in identical fashion for each item, then from Equation 14.19 it follows† that each item amount x_k could be represented by the additive model

$$x_k = \mu_k + \mu_k \beta' + \mu_k \eta_k \quad (14.24)$$

From Equations 14.22 and 14.23 this equation can be written

$$x_k = \mu_k + \mu_k \sum_r \beta'_r + \mu_k \sum_r \eta_{rk} \quad (14.25)$$

†From now on, the approximations in Equations 14.19 through 14.23 will be replaced by equalities.

where β'_r is the relative bias effect due to the r^{th} source of error, and η_{rk} is the relative random effect in item x_k due to the r^{th} source of error. For the observed stratum total $x_i = x = \sum_k x_k$,

$$x = \sum_k \mu_k + (\sum_k \mu_k)(\sum_r \beta'_r) + \sum_r \sum_k \mu_k \eta_{rk} \quad (14.26)$$

or, adopting the same notational convention with respect to sums of the true amounts μ_k as for the observed amount x_k ,

$$x = \mu + \mu \beta' + \sum_r \sum_k \mu_k \eta_{rk} \quad (14.27)$$

However, this ideal state of affairs is seldom present in actual accounting measurements, and the assumption of distinct and separate error effects associated with each item, for each of the m sources of measurement error, is seldom justified. Instead, many parts of the measurement procedures—such as the calibration of standard solutions, NDA instruments, and scales and tanks for volume measurements—are only performed at intervals, and the random effects associated with their replication will apply to a number of items. A similar situation arises when measurement results of element or isotope concentration are based on a limited random sample of items, or a composite sample of aliquots from a number of items. In that case, the random effect of the sampling and analytical procedure is constant for the group of items randomly sampled, or from which aliquots were composited, because a single fixed element or isotopic concentration will be applied to that group of items.† Thus, while the total error in an observed item amount is the total of r independent random effects, the combined error in the stratum total will generally contain many dependent or identical effects. Failure to recognize this dependency can result in a serious underestimation of the error in a stratum total if error propagation is applied directly to supposedly independent errors in item amounts.

Two approaches to the problem are possible. One is to apply the formulas of Section 2.7.1 to the item amounts in the stratum total, taking into account the dependencies between them. This is the usual error propagation approach. The second is to write the observed stratum total as the sum of the true total and a linear combination of independent error effects, and then apply the formulas to these independent contributions to the total error. The coefficients of individual effects will be the number of items containing each effect in the case of absolute effects, and the sum of

†Random effects such as these have been called "short term systematic errors" (Jaech, 1973, p. 81) to indicate that they behave like systematic errors, or biases, for a limited part of a data set.

the item amounts for items containing each effect in the case of relative effects. In either case, a general treatment requires that for each error source the items be separated into classes corresponding to each independent replication of a measurement operation which gives rise to a unique error effect. For example, the items in a stratum could be classified according to the analytical result used to determine the material content of the item. This concept has been used to define the meaning of a "batch" (IAEA, 1982). The same items could be classified according to the scale used to determine bulk weight, in which case the associated "random" effect would be the "bias" in replicate measurements on that particular scale. Note that the random selection of an error effect in this case is with respect to a scale calibration, so that the same set of effects will apply until the scales are recalibrated or replaced.

In general, for each measurement error source r there will be a different classification of the N items in the stratum into a_r batches, where batch $j(r)$ contains $N_{j(r)}$ items whose observed content contains the error effect $\eta_{j(r)}$, $j(r) = 1, \dots, a_r$. The a_r random error effects $\eta_{j(r)}$ apply to the sums $x_{j(r)}$, with expected value $\mu_{j(r)}$ of the measured amounts present in the $N_{j(r)}$ items in batch $j(r)$. This notation is used to indicate that the specific index $j(r)$ is applicable only within the r^{th} error source. Note that $\sum_{j(r)} N_{j(r)} = N$ and $\sum_{j(r)} \mu_{j(r)} = \mu$ for all r . Thus, Equation 14.27 can be rewritten as

$$x = \mu + \mu\beta' + \sum_r \sum_{j(r)} \mu_{j(r)} \eta_{j(r)} \tag{14.28}$$

Note that where a source is completely replicated, it is possible to define $a_r = N$ and $N_{j(r)} = 1$, or $a_r = 1$ and $N_{j(r)} = N$. Whether the replication of a particular measurement operation is considered as performed on every item in a stratum containing one batch, or on N batches of one item each, does not affect the result. The latter is the natural limiting case of the general formulation.

Example 14.1 The situation with respect to the error contribution to the stratum total is best illustrated by a simplified example. Suppose that the stratum under consideration consists of 16 containers of uranium nitrate. Each was weighed separately on the same scale, so there are 16 errors associated with the individual determinations of gross weight. The tare weights of each container before filling were independently determined on a second scale. Four analytical determinations of uranium con-

centration were made on composite samples from each successive set of four containers. Between the third and fourth chemical analysis the analytical procedure was recalibrated, and between the 14th and 15th tare weight determinations the scale was recalibrated. There are six sources of error effects: (1) gross weight determination, (2) tare weight determination, (3) sampling for analysis, (4) analysis, (5) calibration of the analytical procedure, and (6) calibration of the scale used to determine tare weights.

For this example,

$$m = 6$$

$$a_1 = a_2 = a_3 = 16$$

$$a_4 = 4$$

$$a_5 = a_6 = 2$$

$$N_{j(1)} = N_{j(2)} = N_{j(3)} = 1 \text{ for } j = 1, \dots, 16$$

$$N_{j(4)} = 4 \text{ for } j = 1, \dots, 4$$

$$N_{1(5)} = 12, N_{2(5)} = 4$$

$$N_{1(6)} = 14, N_{2(6)} = 2$$

Table 14.2 shows the tabulation of the errors affecting each item for each source. In addition, there will be six relative bias effects $\beta'_1, \dots, \beta'_6$ corresponding to the constant error contribution from each source contained in the 16 observed item amounts in the stratum.

The key consideration is that while there are 12 independent error effects in each estimated item amount, there are not $(12)(16) = 192$ independent effects in the stratum total. The actual number is

$$m + \sum_{r=1}^m a_r$$

or

$$6 + (16 + 16 + 16 + 4 + 2 + 2) = 62$$

TABLE 14.2
Random Effects Applicable to Individual Items

| Item number | True value | Error source | | | | | |
|-------------|------------|----------------|----------------|----------------|---------------|---------------|---------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | μ_1 | $\eta_{1(1)}$ | $\eta_{1(2)}$ | $\eta_{1(3)}$ | $\eta_{1(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 2 | μ_2 | $\eta_{2(1)}$ | $\eta_{2(2)}$ | $\eta_{2(3)}$ | $\eta_{1(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 3 | μ_3 | $\eta_{3(1)}$ | $\eta_{3(2)}$ | $\eta_{3(3)}$ | $\eta_{1(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 4 | μ_4 | $\eta_{4(1)}$ | $\eta_{4(2)}$ | $\eta_{4(3)}$ | $\eta_{1(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 5 | μ_5 | $\eta_{5(1)}$ | $\eta_{5(2)}$ | $\eta_{5(3)}$ | $\eta_{2(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 6 | μ_6 | $\eta_{6(1)}$ | $\eta_{6(2)}$ | $\eta_{6(3)}$ | $\eta_{2(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 7 | μ_7 | $\eta_{7(1)}$ | $\eta_{7(2)}$ | $\eta_{7(3)}$ | $\eta_{2(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 8 | μ_8 | $\eta_{8(1)}$ | $\eta_{8(2)}$ | $\eta_{8(3)}$ | $\eta_{2(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 9 | μ_9 | $\eta_{9(1)}$ | $\eta_{9(2)}$ | $\eta_{9(3)}$ | $\eta_{3(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 10 | μ_{10} | $\eta_{10(1)}$ | $\eta_{10(2)}$ | $\eta_{10(3)}$ | $\eta_{3(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 11 | μ_{11} | $\eta_{11(1)}$ | $\eta_{11(2)}$ | $\eta_{11(3)}$ | $\eta_{3(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 12 | μ_{12} | $\eta_{12(1)}$ | $\eta_{12(2)}$ | $\eta_{12(3)}$ | $\eta_{3(4)}$ | $\eta_{1(5)}$ | $\eta_{1(6)}$ |
| 13 | μ_{13} | $\eta_{13(1)}$ | $\eta_{13(2)}$ | $\eta_{13(3)}$ | $\eta_{4(4)}$ | $\eta_{2(5)}$ | $\eta_{1(6)}$ |
| 14 | μ_{14} | $\eta_{14(1)}$ | $\eta_{14(2)}$ | $\eta_{14(3)}$ | $\eta_{4(4)}$ | $\eta_{2(5)}$ | $\eta_{1(6)}$ |
| 15 | μ_{15} | $\eta_{15(1)}$ | $\eta_{15(2)}$ | $\eta_{15(3)}$ | $\eta_{4(4)}$ | $\eta_{2(5)}$ | $\eta_{2(6)}$ |
| 16 | μ_{16} | $\eta_{16(1)}$ | $\eta_{16(2)}$ | $\eta_{16(3)}$ | $\eta_{4(4)}$ | $\eta_{2(5)}$ | $\eta_{2(6)}$ |

From Equation 14.28, the stratum total x can be written in the form

$$\begin{aligned}
 x = & \mu + \mu\beta'_1 + \mu\beta'_2 + \mu\beta'_3 + \mu\beta'_4 + \mu\beta'_5 + \mu\beta'_6 \\
 & + \mu_1 \eta_{1(1)} + \dots + \mu_{16} \eta_{16(1)} \\
 & + \mu_1 \eta_{1(2)} + \dots + \mu_{16} \eta_{16(2)} \\
 & + \mu_1 \eta_{1(3)} + \dots + \mu_{16} \eta_{16(3)} \\
 & + (\mu_1 + \mu_2 + \mu_3 + \mu_4)\eta_{1(4)} + (\mu_5 + \mu_6 + \mu_7 + \mu_8)\eta_{2(4)} \\
 & + (\mu_9 + \mu_{10} + \mu_{11} + \mu_{12})\eta_{3(4)} + (\mu_{13} + \mu_{14} + \mu_{15} + \mu_{16})\eta_{4(4)} \\
 & + (\mu_1 + \dots + \mu_{12})\eta_{1(5)} + (\mu_{13} + \dots + \mu_{16})\eta_{2(5)} \\
 & + (\mu_1 + \dots + \mu_{14})\eta_{1(6)} + (\mu_{15} + \mu_{16})\eta_{2(6)}
 \end{aligned}$$

The error in the observed total $x = \sum_{k=1}^N x_k$ has been expressed as a linear combination of independent contributions from the measurement operations actually performed.

There is no reason why the transformation of an accounting result from a linear combination of item amounts into a linear combination of unknown true values and relative error effects cannot be extended from stratum totals to more complex combinations of accounting results. The only change is that the coefficients of the individual effects will be more complex combinations of the true item amounts involving more than one stratum or more than one inventory difference component. Such combinations will be considered in Section 14.3.5.

There are cases where the error effects associated with a given batch of item amounts are themselves combinations of a number of effects due to replicate measurement operations. The simplest example is when a measurement on an individual item is replicated, and the mean result applied to the item amount. A more complex example is the situation where a single result, such as a concentration factor, is subject to several sources of error, each of which involves replicated measurement operations not uniquely associated with specific subsets of items. Because the assumption here is that the effects of individual measurement operations are additive, it is sufficient for the treatment of expected variances of stratum totals in the next section to consider that each effect may be the result of b_r replications of the measurement operation for each of the a_r batches.

14.3 EXPECTED VALUES AND VARIANCES OF ACCOUNTING RESULTS

In the preceding section, a linear model was developed expressing the error in a stratum total due to measurement as a linear combination of the errors arising from individual measurement operations. In this section, the results of Section 2.7 will be applied to this model to derive the expected value and variance of the error due to measurement in strata, component totals, and inventory differences from the assumed or estimated means and variances associated with the errors in individual measurements.

14.3.1 Assumed Properties of Error Effects

Each of the a_r error effects $\eta_{j(r)}$ associated with a source of variability r is assumed to be the mean of b_r independent measurement errors with expected value zero and common variance δ_r^2 , so that

$$\text{Var}(\eta_{j(r)}) = E(\eta_{j(r)}^2) = \delta_r^2/b_r \quad (14.29)$$

The expectation is over the replicate measurement operations involved, not the item amounts to which an individual error effect applies. In general, the total number of identically distributed random variables for each source is $a_r b_r$, not N .

Just as with the combination of item amounts into stratum totals, there are computational problems associated with the fact that individual random effects may affect item amounts in more than one stratum, or item amounts in the same stratum in successive inventory differences. In Example 14.1, the effects associated with Source 5 (recalibration of the analytical procedure) will apply to any item amounts in other strata or in subsequent material balance periods for which the element concentrations were determined using the same analytical procedure and the same standardization. Similarly, the calibration error in a particular scale, such as that considered in Source 6 of the example, will apply to all unit amounts weighed on that scale until the scale is recalibrated, leading to an independent replication of the calibration error. As before, failure to correct for the positive covariance between the errors in totals for different strata induced by such infrequent replication of some part of the measurement operation will result in underestimating the variances of accounting results involving the combination of stratum totals. This problem is considered explicitly in Section 14.3.5 and Examples 14.5 and 14.6.

The situation with respect to the combined effect of the biases β_r' is much more complex. One can consider each of these contributions to the total error during a particular accounting period as a random effect which was simply not replicated during the period in question—i.e., the calibration period and accounting period were the same. For example, consider the two scales used to obtain gross and tare weights in Example 14.1. For each scale the routine weighing is one source of measurement error and the calibration of the scale is a second source. The calibration is only repeated periodically, and the calibration of the scale used to determine gross weights was not repeated during the determination of the content of the 16 items in the particular stratum for the particular material balance period. This means that the error effect β_1' in the stratum total due to the gross weight operation is actually the sum of any residual bias effect due to the calibration process—e.g., a bias in a standard—and an error effect associated with the present calibration of the scale used to determine gross weights. If this calibration is considered as a seventh source of uncertainty in the stratum total, then

$$\beta_1' = \beta_7' + \eta_{1(7)}$$

where the single random error effect $\eta_{1(7)}$ has a variance $\delta_7^2 = E(\eta_{1(7)}^2)$, comparable to the variance δ_6^2 of the two effects $\eta_{1(6)}$ and $\eta_{2(6)}$ associated with the calibration of the scale used to determine tare weight, and the effect β_7' is comparable to the effect β_6' . Because calibration or standardization processes are designed to eliminate existing biases in an instrument or procedure, or to establish an unbiased measurement process, the expectation is that the dominant random error of calibration will be a signifi-

cant part of the bias effect β'_1 . In this case, the constant effect associated with source 1 contains an observed value of a random variable whose variance is essentially identical to that of the two replicate random error effects $\eta_{1(6)}$ and $\eta_{2(6)}$ included in Table 14.3 for source 6.

Other sources of measurement variability, such as those listed in Section 14.1.3, contribute in a somewhat different fashion to the constant error effects in a stratum total. It is well known that consistent differences will exist between measurement results obtained by different people at different times under different physical and environmental conditions. At any given time, or over any short period of time, the set of error effects due to such sources may be fixed. For example, the subset of technicians in a given analytical laboratory at a given time will determine a set of error effects uniquely associated with each operator. This set will change only gradually as personnel are replaced, or perhaps as the individuals become more skilled. Specialized instruments, such as a mass spectrometer or a calorimeter, may be operated by the same person over an entire material balance period. The random effect associated with that particular operator may be fixed for all results determined using element concentrations based on calorimetric analysis or for isotopic concentrations based on mass spectrometric analysis. If different operators are assigned to these instruments on different shifts there will be three such error effects, each associated with a different operator or, equivalently, with a different shift. Instrumental readings may vary over time because of the aging of components, and scale readings can vary with temperature and humidity. These changes produce effects in the measurement results that are serially correlated rather than independently replicated on a periodic basis.

Finally, unknown and unrecognized constant effects will remain in the measurement operation even after every attempt has been made to remove them through calibration, standardization, and other measurement control programs. The ideal measurement system minimizes such sources of error, which are usually indistinguishable from the process variations and material losses that the accounting system is designed to discover.

In this and later chapters, values of the variance due to measurement variability in accounting totals will be derived assuming that the individual contributions β'_r to the total bias β' in a stratum total are random effects that arise from specific measurement operations but are constant for all items in the particular set of accounting results under consideration. The assumption that the β'_r are random variables, rather than fixed but unknown constants, can be justified by noting that the measurement process which leads to the β'_r will be repeated many times, leading to different biases each time. It is the conglomerate of such processes that we are interested in evaluating. This assumption is analogous to considering the β'_r simply as another level of additive effects—i.e., as random variables such that $E(\beta') = \sum_r E(\beta'_r) = 0$ and

$$\text{Var}(\beta'_r) = \theta_r^2 \quad (14.30)$$

Because of the specialized nature of these effects, their variances have been designated by a different lower case Greek letter to distinguish the nature of these single effect contributions. More generally, it could be assumed that the m effects β'_r were associated with m additional "sources" of random error, $r = m+1, \dots, 2m$ with $a_{m+1} = a_{m+2} = \dots = a_{2m} = 1$ —i.e., each replicated once during the period in question. This is consistent with the concept that the measurement system applied to individual accounting items is a combination of individual operations replicated with different frequencies designed to produce, insofar as possible, unbiased estimates of accounting totals. Each replication of each operation produces a random contribution to the measurement error whose variance reflects the relative variability associated with the operation.

For the analogous situation of a two-level hierarchical analysis of variance model involving sources within strata and subgroups within sources, the standard analytical procedures would require the assumption that the β'_r effects corresponding to sources were independently and identically distributed. The treatment here does not require this assumption because the intent is only to reflect the contribution of the variance of each random effect to the variance of the stratum total, not to estimate a variance associated with this level of effect. The concern here is only with the process of deriving variances of stratum totals in terms of variances associated with known sources of variability in the measurement operations. The problems involved in estimating the component variances θ_r^2 and δ_r^2 associated with the various sources of measurement variability are considered in Chapter 15.

14.3.2 Expected Value and Variance of Stratum Totals

The expected value and variance of a stratum total can be obtained directly by applying the results given in Section 2.7.1 to the error structure defined in Section 14.2.3, and the expected values and variances of the random error effects from each source of variability defined in Section 14.3.1. The definition of the stratum total x in Equation 14.28, with β' given by Equation 14.22, is a linear function of the form given in Equation 2.121. The random variables are the m random effects β'_r , which are combined additively to obtain β' , and the $\sum_{r=1}^m a_r$ random effects $n_{j(r)}$. All of these effects are assumed to be independent and because their expected values are assumed to be zero, the expected value of the product of any two different effects is zero. The coefficients are the sums $\mu_{j(r)}$ of the true

item amounts associated with each replicated random effect $\eta_{j(r)}$ and the true stratum total μ associated with each random effect β_r not replicated within the stratum. Hence, Equations 2.122 and 2.123 can be used to obtain directly, for the stratum total x ,

$$E(x) = \mu + \mu E(\beta') + \sum_r \sum_{j(r)} \mu_{j(r)} E(\eta_{j(r)}) \quad (14.31)$$

$$\text{Var}(x) = \mu^2 \sum_r \theta_r^2 + \sum_r (\delta_r^2/b_r) (\sum_{j(r)} \mu_{j(r)}^2) \quad (14.32)$$

Equation 14.32 is valid only under the assumption of identical distributions for the effects $\eta_{j(r)}$ within each source of measurement variability with variances given by Equation 14.29.

Assuming, as in the previous section, that $E(\beta') = \sum E(\beta'_r) = 0$ and that $E(\eta_{j(r)}) = 0$ for all r , it follows from Equation 14.31 that $E(x) = \mu$. These assumptions are consistent with the recommendation in a study of deficiencies in the statistical analysis of inventory differences (Byers et al., 1983) that insofar as the measurement process is concerned, "an attempt should be made to estimate and correct for all potentially significant biases. To the extent that this is accomplished, the estimator of inventory difference will be free of 'fixed' errors or biases."

There has been some controversy over these assumptions, but they seem reasonable with respect to an effective and well-controlled measurement system. There are two different problems involved. The first is the existence of strata which cannot be measured directly, such as material in processes at the time of an inventory. The estimation of these quantities from historical data or subsequent measurement can be considered simply as another source of measurement error, even though the estimation process may be indirect and highly variable. The second problem is the existence of unknown and unrecognized process losses or measurement biases. These represent a lack of complete control of the accounting and measurement process, and their discovery and elimination is a primary concern of the material accounting process. Their existence detracts from the ability to discover accidental losses or deliberate diversions. Their effect on the interpretation of inventory differences is discussed further in Chapter 16.

Example 14.2 Continuing Example 14.1, assume:

$$\begin{aligned} E(\beta'_r) &= 0 & r &= 1, \dots, 6 \\ \text{Var}(\beta'_r) &= E(\beta_r'^2) = \theta_r^2 & r &= 1, \dots, 6 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(1)}) &= 0 & j &= 1, \dots, 16 \\ \text{Var}(\eta_{j(1)}) &= E(\eta_{j(1)}^2) = \delta_1^2 & j &= 1, \dots, 16 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(2)}) &= 0 & j &= 1, \dots, 16 \\ \text{Var}(\eta_{j(2)}) &= E(\eta_{j(2)}^2) = \delta_2^2 & j &= 1, \dots, 16 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(3)}) &= 0 & j &= 1, \dots, 16 \\ \text{Var}(\eta_{j(3)}) &= E(\eta_{j(3)}^2) = \delta_3^2 & j &= 1, \dots, 16 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(4)}) &= 0 & j &= 1, \dots, 4 \\ \text{Var}(\eta_{j(4)}) &= E(\eta_{j(4)}^2) = \delta_4^2 & j &= 1, \dots, 4 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(5)}) &= 0 & j &= 1, 2 \\ \text{Var}(\eta_{j(5)}) &= E(\eta_{j(5)}^2) = \delta_5^2 & j &= 1, 2 \end{aligned}$$

$$\begin{aligned} E(\eta_{j(6)}) &= 0 & j &= 1, 2 \\ \text{Var}(\eta_{j(6)}) &= E(\eta_{j(6)}^2) = \delta_6^2 & j &= 1, 2 \end{aligned}$$

The coefficients of the random effects in the expression for the stratum total are as follows:

$$\begin{aligned} \mu &= \sum_{k=1}^{16} \mu_k \\ \mu_{j(1)} &= \mu_{j(2)} = \mu_{j(3)} = \mu_k, \quad k = 1, \dots, 16 \\ \mu_{1(4)} &= \mu_1 + \mu_2 + \mu_3 + \mu_4 \\ \mu_{2(4)} &= \mu_5 + \mu_6 + \mu_7 + \mu_8 \\ \mu_{3(4)} &= \mu_9 + \mu_{10} + \mu_{11} + \mu_{12} \\ \mu_{4(4)} &= \mu_{13} + \mu_{14} + \mu_{15} + \mu_{16} \\ \mu_{1(5)} &= \sum_{k=1}^{12} \mu_k \\ \mu_{2(5)} &= \mu_{13} + \mu_{14} + \mu_{15} + \mu_{16} \\ \mu_{1(6)} &= \sum_{k=1}^{14} \mu_k \\ \mu_{2(6)} &= \mu_{15} + \mu_{16} \end{aligned}$$

Substituting these expected values and variances in Equation 14.32 yields

$$\begin{aligned}
 \text{Var}(x) = & \mu^2 \sum_{r=1}^6 \theta_r^2 \\
 & + (\delta_1^2 + \delta_2^2 + \delta_3^2) \sum_{k=1}^{16} \mu_k^2 \\
 & + \delta_4^2 [(\mu_1 + \mu_2 + \mu_3 + \mu_4)^2 + (\mu_5 + \mu_6 + \mu_7 + \mu_8)^2] \\
 & + (\mu_9 + \mu_{10} + \mu_{11} + \mu_{12})^2 + (\mu_{13} + \mu_{14} + \mu_{15} + \mu_{16})^2 \\
 & + \delta_5^2 [(\mu_1 + \dots + \mu_{12})^2 + (\mu_{13} + \dots + \mu_{16})^2] \\
 & + \delta_6^2 [(\mu_1 + \dots + \mu_{14})^2 + (\mu_{15} + \mu_{16})^2]
 \end{aligned}$$

Equation 14.32 can be written in the form

$$\text{Var}(x) = \sum_r [\mu^2 \theta_r^2 + (\sum_{j(r)} \mu_{j(r)}^2)(\delta_r^2/b_r)] \quad (14.33)$$

When written in this form it can be seen that the contribution to the variance from each source of variability is an individual term, analogous to the contribution of each class to the variance of the completely weighted overall mean in the analysis of variance model for nested random effects and unequal numbers considered in Section 5.3.1.2. If $\bar{\mu} = N\mu$ and each source of measurement variability has the same absolute variance components $\sigma_\alpha^2 = \mu^2 \theta_r^2$ and $\sigma_\beta^2 = \mu^2 \delta_r^2$ at the first and second levels of the hierarchy (which correspond to biases and individual measurement effects, respectively), then Equation 14.33 reduces to the form of Equation 5.136 with $\sigma^2 = 0$, $i = r$, $n_{ij} = N_{j(r)}$, $n_i = N$, $n_{..} = mN$, $a = m$, and $b_i = a_r b_r$. As noted previously, when the concern is with estimating individual variance components associated with a given source of variability, the emphasis is on classes of effects with constant variance. When computing the expected variability in a stratum total, all sources of variability should be included without regard to their degree of replication and consequent estimability from the accounting results.

14.3.3 Calculating the Variance of Stratum Totals

Equations 14.32 and 14.33 contain two sets of unknown parameters, the true content of the individual items and the true variances of the ran-

dom error effects contributing to the variability in the estimated amounts of nuclear material. Knowledge of the unknown true amounts is required only because the random error effects are assumed to be proportional to the true content. However, assuming identically distributed absolute effects would require knowing not only the true values μ_k of the item amounts, but also the values of the individual measurements which enter multiplicatively into their computation.

Consider first the bias in the estimation procedure introduced by the use of measured amounts instead of true amounts, assuming the relative variance components θ_r^2 and δ_r^2 are known. The variance computed using measured rather than true item amounts will be designated by $\text{Var}'(x)$ rather than $\text{V}\hat{\text{a}}\text{r}(x)$. This distinction is necessary because only the item amounts, and not the relative error variance components, are estimated. If in Equation 14.32 the quantities μ and $\mu_{j(r)}$ are replaced by x and $x_{j(r)}$, respectively, the variance actually computed is

$$\text{Var}'(x) = x^2 \sum_r \theta_r^2 + \sum_r (\delta_r^2/b_r) (\sum_{j(r)} x_{j(r)}^2)$$

Because for any random variable, $E(X^2) = [E(X)]^2 + \text{Var}(X)$, it follows that

$$E[\text{Var}'(x)] = \text{Var}(x) + \text{Var}(x) \sum_r \theta_r^2 + \sum_r (\delta_r^2/b_r) \sum_{j(r)} \text{Var}(x_{j(r)})$$

For all r , $\sum_{j(r)} x_{j(r)} = x$. Therefore, $\sum_{j(r)} \text{Var}(x_{j(r)}) \leq \text{Var}(x)$, because there must be at least one other source of variation with effects common to two or more of the measurement totals $x_{j(r)}$ unless all other sources are completely nested within r , in which case the equality holds for that particular r . Hence,

$$E[\text{Var}'(x)] \leq \text{Var}(x) + \text{Var}(x) [\sum_r \theta_r^2 + \sum_r \delta_r^2/b_r]$$

Thus, on the average, using observed rather than true values to compute the variance of the total will inflate the estimate by a relative amount less than the sum of the relative variances for each individual source. Because these relative variances are usually quite small (e.g., a relative standard deviation of 1% or $\delta = .01$ corresponds to a variance $\delta^2 = .0001$), this source of error is usually quite small.

The situation with respect to the replacement of the variances δ_r^2 and θ_r^2 by estimates is much more complex. It was noted in the previous section that estimation of the variance components associated with specific error sources is possible only if (1) the measurements contain replicate effects arising from the error source whose variability is being estimated;

and (2) the measurements are on items with known true values or the same unknown true value. These conditions do not normally hold for routine production measurements. Not only is the true content not known, but replicate measurements are not usually made on production items. Information concerning measurement variability can be obtained from repeated measurements on individual production items for those sources usually replicated in such measurements. Otherwise, estimates must be obtained from planned replication on production items of certain classes of error effects, or from the calibration, standardization, and monitoring procedures used to establish and control the measurement system. There are two basic concerns over the use of the latter to compute the variability in the accounting results. The first is whether the variance components present in standardization and control procedures are comparable to those present in actual accounting measurements. The second is the difficulty in establishing the quality of the estimates, which may be based on a wide variety of both contemporary and historical data analyzed by a number of different direct and indirect statistical techniques.

The wide variation in suggested approaches to the calculation of the variability in accounting results can be described in terms of the complexity of the assumed error sources and the degree of simplification in the form of the error structure made for computational purposes. These elements are not entirely independent, as simplification in the assumed error sources may be a means of simplifying the computations required. The treatment in the preceding section allows for as many sources of measurement variability as can be identified, and a completely general association of the item amounts with each random error effect due to each source. However, the assumption that each error source produces random effects proportional to the true values of the item amounts to which they apply is relatively restrictive, as was discussed in Section 14.2.1. A more general treatment would allow for a more complex characterization of the errors in an estimate of the item amount that involved both the true amount present and the actual measurement procedure. Franklin (1984) points out the general form required and illustrates its application to an NDA determination involving the combination of several counting measurements, each of which has a variance proportional to the total count. Such treatments, of course, presume a detailed knowledge of the specific measurement procedure. The distinction is between detailed treatments of the error structure that are possible for specific plants with well-defined processes and measurement procedures, and generic treatments that have been used to characterize measurement systems generally and to simplify the treatment of individual facilities. In addition to the example given in Appendix 14A, a good example of the complete treatment of the measurement error structure in a specific facility is given in NUREG/CR-3221 (Stirpe, Goldman, and Gutmacher, 1983). This treatment includes not only the system design

and error structure for the accounting process, but also the evaluative procedures to be considered in Chapter 16.

As noted in Section 14.1.3, most generic treatments define three parts of the measurement process—bulk measurement method, sampling, and analysis—as primary sources of measurement error in a specific material balance. Where accounting is for an isotope rather than an element, sampling and analysis of the isotope must be included for a total of five primary sources. Within each primary source, most generic treatments (Jaech, 1973) distinguish three classifications of the errors according to the degree of replication of the particular measurement operation: “random” effects replicated independently for each item; “short-term systematic” effects, which are replicated periodically and have a common effect on groups of items within a stratum; and “long-term systematic” effects common to all items in a stratum. All of these effects can be treated as observations of random variables in the statistical sense, and the above classification is just one way of recognizing the disparity in the replication of the various operations or procedures contributing to the measurement error. Jaech (1973, p. 185) notes that although he defines only 15 types of error components (random, short-term systematic, and long-term systematic contributions from each of five primary parts of the measurement process), “there are generally many more separately identified components in a given operation.” Similarly, although the AMASS procedures (Messinger, Lumb, and Tingey, 1981; and Lumb, Messinger, and Tingey, 1983) consider only nine generic sources (three primary measurement operations and three levels of replication), later applications of this generic structure consider for a specific example the multiple sources of error within each generic source (Tingey and Lumb, 1984).

Given that the assumption of constant variance of the relative error effects from a given source is acceptable, there is no theoretical difficulty in treating the specific individual contributions to the variance of a stratum total from an arbitrarily large number of sources. The reasons for the restricted classifications of the previous paragraph are primarily the practical difficulties of (1) computing the totals of those specific item amounts contained in the accounting result to which a specific error effect applies; and (2) obtaining individual estimates of the variances of the random effects from each source. Considering the latter first, the question is again one of generic or specialized treatment of a particular application. For a given facility and process, experiments can be defined and carried out to obtain individual estimates of the applicable variances. Chapter 15 considers a selection of procedures of this type. However, general treatments of measurement capability used in the design of evaluative procedures and safeguards approaches are based on historical compilations of error variances for a limited set of generic sources and material types and known capabilities of analytical procedures (Reilly and Evans, 1977; Hakkila et al., 1980; Rogers, 1983; DeBieve, 1984).

The primary simplification of the variance computation procedure has been with respect to the totals $\mu_{j(r)}$ required to compute the coefficients $\Sigma_{j(r)} \mu_{j(r)}^2$ of the individual variance components in Equations 14.32 and 14.33. Jaech (1973) gives rules for the computation of these coefficients ("C" coefficients, in his terminology) for the 15 generic error sources noted previously. The problem is that these computations require a record system that makes it possible to associate the item amount for each item in a particular accounting total with the specific measurement operations, or sources of error, affecting that item. The computation can be greatly simplified by assuming only error effects that are constant for all items or unique to each item. In this case it is only necessary to know the sources of measurement error, or "error path" (Argentisi, Casilli, and Franklin, 1979), associated with each item and the "random" and "systematic" error variances defined by Jaech (1973) associated with each source. This assumption eliminates the need for keeping track of the item amounts (or, equivalently, the absolute contribution to the total error) associated with each replication of a specific measurement operation. Other approaches attempt to obtain the necessary information by associating the definition of certain reporting units, usually called batches, with the replication of one or several parts of the measurement operation. This use of terminology normally associated with process operations to quantities determined by measurement operations is natural in some instances, but on other occasions may be quite artificial.

The most simplified computational approach appearing in the literature (Messinger, Lumb, and Tingey, 1981; IAEA, 1982) is based on assumptions similar to those used in Section 14.3.2 to show the analogy between the variance of an observed stratum total and the expected mean square for the mean in an analysis of variance. The two assumptions are: (1) that all items have the same true value—i.e., $\mu_k = \bar{\mu} = \mu/N$ for all N items in the stratum; and (2) that all of the a_r "batches" of items associated with unique error effects contain the same number of items $N_r = N/a_r$. Under these assumptions $\mu_{j(r)} = \mu/a_r$, and Equation 14.32 reduces to

$$\text{Var}(x) = \mu^2 \left[\sum_r (\theta_r^2 + \delta_r^2/a_r b_r) \right] \quad (14.34)$$

This computational formulation is convenient because it is necessary only to know the total amount of material μ in the stratum, not in any individual item. The relative variance applicable to this total (the term in the square brackets in Equation 14.34) is quite generally obtained by summing the relative variance divided by the number of replicate measurements over all sources of measurement variability.

Numerous computational packages of varying degrees of sophistication have been developed to carry out the error propagation required to compute $\text{Var}(x)$. One of the more complex is NUMSAS, developed for use in

the European community (Argentisi, Casilli, and Franklin, 1979). Others have been developed by various organizations in the United States, such as Brookhaven and Los Alamos National Laboratories.

Example 14.3 Examples 3.3(a) and 3.5(a) of Volume 3, Part F of the Safeguards Technical Manual (IAEA, 1982) consider a simplified material balance for a fuel fabrication facility. The first of 7 strata consists of 12,000 containers of UO_2 powder, each nominally containing 20 kg of uranium. These are divided into 80 batches of 150 containers each. The amount of UO_2 is determined by weighing each individual container, but the concentration factor used to convert bulk weight to uranium content is determined by obtaining five samples at random from each batch and analyzing each sample for concentration factor. The error structure assumes that there are six variance components: three associated with the individual relative error effects due to each bulk measurement, each sample taken, and each chemical analysis; and three associated with constant relative effects over the entire stratum associated with each of these operations. The values of the δ_r and θ_r , and $r = 1, 2, 3$ are given as

$$\delta_1 = 0.000658 \quad \theta_1 = 0.000439$$

$$\delta_2 = 0.000531 \quad \theta_2 = 0$$

$$\delta_3 = 0.000433 \quad \theta_3 = 0.000571$$

From the description of the measurement operation,

$$\begin{array}{lll} a_1 = 12,000 & b_1 = 1 & N_1 = 1 \\ a_2 = 80 & b_2 = 5 & N_2 = 150 \\ a_3 = 80 & b_3 = 5 & N_3 = 150 \end{array}$$

From Equation 14.34,

$$\begin{aligned} \text{Var}(x) &= (240,000)^2 [(0.000439)^2 + (0.000571)^2 \\ &\quad + (0.000658)^2/12,000 + (0.000531)^2/400 \\ &\quad + (0.000433)^2/400] \\ &= (240,000)^2 (5.199717 \times 10^{-7}) \\ &= 29,950.37 \text{ (kg)}^2 \end{aligned}$$

which corresponds to a standard deviation of 173.06 kg of uranium.

Note that in this example, the error effects in the individual operations contribute only 1.2097×10^{-9} to the total relative variance, correspond-

ing to an absolute variance in the measured amount of $(240,000)^2 \times (1.2097 \times 10^{-9}) = 69.678 \text{ (kg)}^2$, or a standard deviation of 8.35 kg. This large difference emphasizes the importance of the assumptions concerning the nature of the constant effects in determining the variance in an accounting total.

The same result is obtained by the following approach, which follows from the general result given in Equation 14.32. There are 2 error effects associated with the entire stratum, 80 from each of 2 sources associated with batches of 150 items, and 12,000 associated with each individual item. The two sources associated with batches are each replicated five times. Thus

$$\begin{aligned} \text{Var}(x) &= (240,000)^2 [(0.000439)^2 + (0.000571)^2] \\ &\quad + (0.000658)^2 (12,000)(20)^2 \\ &\quad + [(0.000531)^2 / 5](80)(3000)^2 \\ &\quad + [(0.000433)^2 / 5](80)(3000)^2 \\ &= 29,950.37 \text{ (kg)}^2 \end{aligned}$$

because $\sum_{j(r)} \mu_{j(r)}^2$ reduces to 80 (3000)² for the case of 80 equal batches of 3000 kg each.

Because by definition strata usually consist of similar items, the more crucial of the two assumptions required to obtain Equation 14.34 is the assumption of equal numbers of items per batch or more generally, equal numbers of items per replication of the measurement operation associated with an error source. Procedures for replication of measurement operations on individual items or for combining items into batches for measurement purposes, such as the compositing of samples from four items for analysis, usually satisfy the requirement for equal numbers of items per replication and replications per item. However, error sources associated with processing operations, material type, measuring equipment, standardization, and calibration frequently involve replication which is quite nonuniform with respect to both time and material throughout. The actual error associated with the assumption of equal batch sizes is easily determined because the approximation involved is the replacement of $\sum_{j(r)} \mu_{j(r)}^2$ by μ^2/a_r . The difference between these two quantities is $\sum_{j(r)} (\mu_{j(r)} - \bar{\mu}_r)^2$, where $\bar{\mu}_r = \mu/a_r$. Thus, for a given error source, the underestimation of the variance in the stratum total due to the assumption of equal batch sizes is directly proportional to the variability in the batch sizes. In practice, this difference is usually small compared with μ^2/a_r . Differences in the number of items per batch and the true amount per item both contribute to the variability in batch content.

Example 14.4 Suppose that in Example 14.3 a routine scale calibration program was instituted to eliminate the rather large contribution to the variance of the observed stratum total due to the magnitude of the variance component associated with the single constant error effect in the determination of the bulk amount of UO_2 in each container. As a result, the variance component $\theta_1^2 = (0.000439)^2$ was divided into two components $\theta_1'^2$ and δ_4^2 , $\theta_1'^2 = (0.000110)^2$ and $\delta_4^2 = (0.000425)^2$. The variance component δ_4^2 is associated with the individual biases in each successive calibration, while $\theta_1'^2 (= \theta_1^2 - \delta_4^2)$ is the variance component associated with the single constant effect corresponding to the remaining overall bias in the combined calibration and measurement process. In cases like this it is not necessary (and often not possible) to distinguish between the variance component $\theta_1'^2$ associated with a constant error effect in the bulk measurement operation, and the variance component θ_4^2 associated with a constant error effect in the calibration process. The single component $\theta_1'^2 = \theta_1^2 + \theta_4^2$ is sufficient. A similar situation holds for the errors of sampling and chemical analysis associated with the measurement of element concentration. This was the basis for the assumption of $\theta_2 = 0$ in Example 14.3.

For the material balance period of Example 14.3, suppose that 11 calibrations were carried out, so that the 12,000 weighings involved 12 different error effects due to calibration. Suppose also that the actual calibration schedule and the number of weighings associated with each of the 12 periods were as follows:

| Period | Calibration following weighing number | Number of items weighed |
|--------|---------------------------------------|-------------------------|
| 1 | 850 | 850 |
| 2 | 2,250 | 1,400 |
| 3 | 3,425 | 1,175 |
| 4 | 3,910 | 485 |
| 5 | 5,465 | 1,555 |
| 6 | 6,125 | 660 |
| 7 | 7,100 | 975 |
| 8 | 8,620 | 1,520 |
| 9 | 9,275 | 655 |
| 10 | 9,980 | 705 |
| 11 | 11,225 | 1,245 |
| 12 | 12,000+ | 775 |

Because each container weighs a nominal 20 kg, the coefficient associated with the variability δ_4^2 in the calibration process is

$$\begin{aligned}\sum_{j(4)} \mu_{j(4)}^2 &= (20)^2[(850)^2 + (1400)^2 + \dots + (1245)^2 + (775)^2] \\ &= (20)^2(13,489,700) \\ &= (73,456.65)^2 (\text{kg})^2\end{aligned}$$

The combined contribution of the single residual bias effect and the recalibration process to the variance is

$$\begin{aligned}\mu^2\theta_1^2 + \delta_4^2 \sum_{j(4)} \mu_{j(4)}^2 &= (240,000)^2(0.000110)^2 \\ &\quad + (73,456.65)^2(0.000425)^2 \\ &= 1671.59 (\text{kg})^2\end{aligned}$$

This compares to the previous contribution of

$$\begin{aligned}\mu^2\theta_1^2 &= (240,000)^2(0.000439)^2 \\ &= 11,100.73 (\text{kg})^2\end{aligned}$$

associated with the bias effect in the uncalibrated weighing operation.

If the approximation in Equation 14.34 had been used, the contribution of the $a_r = 12$ single replications of the recalibration process would have been computed as

$$\mu^2\delta_4^2/a_4b_4 = [(240,000)^2/12](0.000425)^2 (\text{kg})^2$$

The exact coefficient $(73,465.65)^2$ has been replaced by the approximate coefficient $[(240,000)^2/12] = (69,282.03)^2$. This reduction is relatively small, even though there is considerable variability in the number of weighings associated with each calibration. In extreme cases, a single restandardization or recalibration occurs near the beginning or end of a period. The underestimation of the coefficient, and hence the contribution to the variability of the error source, approaches a factor of one-half under these circumstances.

The assumptions made and the degree of approximation tolerable in a given situation are highly dependent on the specific accounting situation and the purpose of the data processing and analysis. Control procedures

involving frequent material balances across specific unit processes, where the measurement operations are well defined and the accounting and measurement control data are immediately available, may justify a precise definition of the variance and the experiments necessary to estimate the variance components due to all identifiable specific error sources. At the other extreme, when the intent is to examine possible approaches to the evaluation of reported inventory differences for major facilities at intervals of 3 to 6 months, such precise definitions of the sources of variability may not be necessary. In the latter case, "historical" data are normally used to determine the individual variance components contributing to the inventory difference, and the process is generally described in generic terms, so that approximate procedures may be completely adequate.

14.3.4 Variance of the Variance Estimator

In the previous section, the concern was with the adequacy of the model used to relate the variance of a stratum total to the component variances. The validity of this relationship is determined primarily by the exactness of the assumed error structure and the approximations used in modeling. This section is concerned with the contribution of the variability in the estimates of the individual variance components to the variability in the estimated total variance.

Two approaches can be used to compute the variance of the estimator of the total variance. If in Equation 14.32 for $\sigma_x^2 = \text{Var}(x)$ the true variances θ_r^2 and δ_r^2 are replaced by independent unbiased estimates, Equation 2.123 of Section 2.7.1 can be applied to the resulting linear combination of random variables if the relatively small effect of replacing μ and $\mu_{j(r)}$ by their estimates is neglected. It follows that the estimator

$$\hat{\sigma}_x^2 = \mu^2 \sum_r \hat{\theta}_r^2 + \sum_r (\hat{\delta}_r^2 / b_r) (\sum_{j(r)} \mu_{j(r)}^2) \quad (14.35)$$

has variance

$$\text{Var}(\hat{\sigma}_x^2) = (\mu^2)^2 \sum_r \text{Var}(\hat{\theta}_r^2) + \sum_r (\sum_{j(r)} \mu_{j(r)}^2 / b_r)^2 \text{Var}(\hat{\delta}_r^2) \quad (14.36)$$

From the discussion of the distribution of an estimator S^2 of σ^2 with ν degrees of freedom in Section 3.5.1, it follows that for sampling from a normal distribution

$$\text{Var}(S^2) = \frac{2\sigma^4}{\nu} \quad (14.37)$$

Thus if the estimates $\hat{\theta}_r^2$ and $\hat{\delta}_r^2$ are based on degrees of freedom ν_r' and ν_r , $r = 1, \dots, m$, respectively,

$$\text{Var}(\hat{\sigma}_x^2) = 2[(\mu^2)^2 \sum_r (\theta_r^2)^2 / \nu_r' + \sum_r (\sum_{j(r)} \mu_{j(r)}^2)^2 (\delta_r^2)^2 / \nu_r] \quad (14.38)$$

This is simply twice the sum of the squares of the absolute contribution of each source of variability to the total variance divided by the appropriate degrees of freedom and is completely analogous to the usual formula for the variance of a sum of independent variances.

To actually estimate $\text{Var}(\hat{\sigma}_x^2)$ using Equation 14.38, it is again necessary to replace the true variance components by their estimates. The nature of the variability in the estimate of the variance resulting from this procedure is best approximated by the method introduced in Sections 5.3.1.1 and 5.3.1.2 to determine the approximate degrees of freedom in a combination of variance estimators. This approach derives an approximate degrees of freedom for the linear combination of variance estimators which can then be used in conjunction with Equation 14.37 to obtain the variance of the variance estimate. Moreover, this procedure makes it possible to perform approximate tests of hypotheses concerning the variance in accounting results. From Equation 3.54, the general formula for the approximate degrees of freedom for an estimate $\hat{\theta}^2$ of the variance θ^2 , where $\hat{\theta}^2 = \sum_i a_i \hat{\theta}_i^2$ and the $\hat{\theta}_i^2$ are independent estimates of θ_i^2 distributed as chi-square with ν_i degrees of freedom, is

$$\nu = \hat{\theta}^4 / \sum_i (a_i \hat{\theta}_i^2)^2 / \nu_i \quad (14.39)$$

If this general formula is applied to Equation 14.35, it follows that

$$\nu = (\hat{\sigma}_x^2)^2 / [\sum_r (\mu^2 \hat{\theta}_r^2)^2 / \nu_r' + \sum_r (\hat{\delta}_r^2 \sum_{j(r)} \mu_{j(r)}^2)^2 / \nu_r] \quad (14.40)$$

This can be approximated by

$$\nu \cong 2(\hat{\sigma}_x^2)^2 / \hat{\text{Vâr}} \hat{\sigma}_x^2 \quad (14.41)$$

where $\hat{\text{Vâr}}(\hat{\sigma}_x^2)$ is obtained by replacing θ_r^2 and δ_r^2 by $\hat{\theta}_r^2$ and $\hat{\delta}_r^2$, respectively, in Equation 14.38.

The primary difficulty in using a chi-square distribution with degrees of freedom ν , given by Equation 14.40 to approximate the distribution of $\hat{\text{Vâr}}(\hat{\sigma}_x^2)$, is the frequent lack of information concerning the degrees of freedom ν_r and ν_r' in the individual variance estimates. Where the estimates of

the component is based on long experience with a particular instrument, such as a scale or an NDA device, the assumption of a known variance component (infinite degrees of freedom, zero contribution to the variance of the estimator) may be warranted. On the other hand, it is frequently apparent from the nature of the available data on particular variance components that the assumption that they are known is not warranted. Whether or not the actual variance (or equivalent degrees of freedom) can be calculated, Equations 14.39 and 14.40 show that variance component estimates based on small degrees of freedom may contribute a disproportionate amount of the variability in the estimates of the variance of stratum totals. In particular, estimating the variance of a single constant error effect from the deviation of an overall mean from a standard value can be highly variable. This can be seen by looking at the 5% and 95% critical values of χ^2 for 1 degree of freedom, whose reciprocals are the factors to be applied to the squared deviation to determine the 90% confidence limits on the true variance component (Section 3.5.1). The principal problem arises from observed means that, by chance, fall close to the standard. This possibility leads to the expectation that 5% of the time the true variance component will be underestimated by a factor of more than 254 when the estimate is based on 1 degree of freedom. The situation is compounded by the fact that the contributions associated with single random error effects, whose absolute contributions to the total error are directly proportional to the stratum total, tend to be the dominant contributions to the total variance. Thus, it is particularly important that sound historical experience be accumulated with respect to the expected variability due to this type of effect.

14.3.5 Combining Variances for Strata and Components

The procedures for individual strata developed in the preceding section can be extended to any combination of strata or, more generally, any linear combination of individual item amounts. In particular, the extension is easily made to major components of the inventory difference, such as beginning and ending inventories; to the inventory difference itself; and to the combination of similar strata, such as product shipments or measured discards, over several material balance periods. The computation of the variance estimator for a combination of several strata is a simple reapplication of the equation for the variance of a linear combination of random variables to the results for each stratum obtained using the procedures developed in the previous section. The presence of common error effects induces a covariance between the estimators of stratum totals which must be taken into account. It is also possible that two components of the same inventory difference, or the same component of successive inventory differ-

ences, may contain identical item amounts. For example, an item may be both received and carried in the ending inventory at the same value, or may be carried at the same measured value in several successive inventories. Because this corresponds to a complete rather than partial replication of the individual error effects corresponding to two item amounts, the error in these item amounts will not contribute to the error in the inventory difference. This problem can be eliminated by redefining the strata or the components so as to eliminate the identical item amounts.

Generally, the estimate of an inventory difference, or a component thereof, is a linear combination of l stratum totals x_i of N_i items. For the inventory difference the general form is

$$ID = \sum_{i=1}^l c_i x_i \quad (14.42)$$

where the c_i are either $+1$ or -1 depending on the sign with which a particular stratum enters the material balance equation. For a single component total x the c_i are either all $+1$ or all -1 , so that when combining stratum totals for the subset of strata within a component, the formula for the component total is simply

$$x = \sum_i x_i \quad (14.43)$$

In this case, the total number of different (but possibly not independent) estimates of item amounts contained in the total is $N = \sum_i N_i$, where N_i is the number of items in the stratum.

The item amounts contained in the inventory difference or a component total may be affected by some or all of the m sources of variability. A generalized treatment is necessary because, while these sources of variability need not affect all measurements, they are not always unique to a stratum. The error effects are determined by the measurement operations used to obtain the item amounts, not by the use of the item amounts to obtain accounting results. A completely general treatment could also consider the possibility of an error source affecting only some of the elements in a stratum. This was not done in the preceding section because it is inconsistent with the definition of strata in the material accounting sense (Franklin, 1984), and can always be handled by redefining strata. Only the presence of a given error effect in two or more strata will be examined in the following development.

Consider first the situation in which not all of the N item amounts involved in a component contain a contribution from each error source. Substituting the expression given by Equation 14.27 for each stratum total x_i into Equation 14.43, the component total is given by

$$x = \sum_i \mu_i + \sum_i \mu_i \beta'_i + \sum_r \sum_{ik} \mu_{ik} \eta_{rik} \quad (14.44)$$

where $\mu_i = \sum_k \mu_{ik}$ is the true material content of the i^{th} stratum. For each item amount x_{ik} not affected by source r , $\eta_{rik} = 0$. Hence, only some of the $N = \sum_i N_i$ item amounts in the component total will be included in one of the $j(r) = 1, \dots, a_r$ subsets associated with a unique error effect $\eta_{j(r)}$ due to error source r . If $N_{ij(r)}$ is the number of item amounts in stratum i containing an error effect j from source r , then in general $\sum_i N_{ij(r)} = N_{j(r)}$, the total number of item amounts in the component containing the j^{th} effect from the r^{th} source. Because it has been assumed that a given source of variability will affect all item amounts in a given stratum, $\sum_{j(r)} N_{ij(r)} = N_i$. The effects of β'_r and $\eta_{j(r)}$ associated with error source r appear only in some subset $i(r) = 1, \dots, \ell_r$ of the ℓ strata in the component, and $\sum_{i(r)} N_i = \sum_{j(r)} N_{j(r)}$. Based on these definitions, the extension of Equation 14.28 to a component total becomes

$$\begin{aligned} x &= \mu + \sum_r \sum_{i(r)} \mu_i \beta'_r + \sum_r \sum_{j(r)} \eta_{j(r)} \sum_{i(r)} \mu_{ij(r)} \\ &= \mu + \sum_r \mu_{(r)} \beta'_r + \sum_r \sum_{j(r)} \eta_{j(r)} \mu_{j(r)} \end{aligned} \quad (14.45)$$

where

$$\begin{aligned} \mu &= \sum_i \mu_i \\ \mu_{(r)} &= \sum_{i(r)} \mu_i \\ \mu_{j(r)} &= \sum_{i(r)} \mu_{ij(r)} \end{aligned} \quad (14.46)$$

and $\mu_{ij(r)}$ is the sum of the expected values of the item amounts in the i^{th} stratum containing the error effect $\eta_{j(r)}$. The only differences between this formulation for a component and that for a single stratum are that in general, the coefficients $\mu_{(r)}$ of the constant error effects β'_r will vary with r , and that $\sum_{j(r)} \mu_{j(r)}$ is not generally equal to μ but to $\mu_{(r)} \leq \mu$.

If, as in the case of a single stratum, Equation 2.123 is applied to Equation 14.45,

$$\text{Var}(x) = \sum_r [\mu_{(r)}^2 \theta_r^2 + \delta_r^2 / b_r \sum_{j(r)} \mu_{j(r)}^2] \quad (14.47)$$

If $\text{Var}(x)$ had been obtained simply by summing Equation 14.33 for a single stratum over the ℓ strata, the result would have been

$$\text{Var}(x) = \sum_r \sum_{i(r)} \mu_i^2 \theta_r^2 + \sum_{j(r)} \delta_r^2 / b_r (\sum_{i(r)} \mu_{ij(r)}^2) \quad (14.48)$$

The difference between Equations 14.47 and 14.48 is the cross product terms of the form $\mu_{i,(r)}\mu_{i,(r)}\theta_r^2$ and $\mu_{ij,(r)}\mu_{ij,(r)}\delta_r^2/b_r$, which are the covariances between the individual strata. For example,

$$\begin{aligned}\text{Cov}[(\mu_{i_1,(r)}\beta_r'),(\mu_{i_2,(r)}\beta_r')] &= E[(\mu_{i_1,(r)}\beta_r')(\mu_{i_2,(r)}\beta_r')] \\ &= \mu_{i_1,(r)}\mu_{i_2,(r)}E(\beta_r'^2) \\ &= \mu_{i_1,(r)}\mu_{i_2,(r)}\theta_r^2\end{aligned}\quad (14.49)$$

Thus, there are two possible ways to compute component variances when the same error effects are present in more than one stratum. The first is to sum the items involving each specific error effect over all strata in the component. The second is to add all of the positive covariances of the form of Equation 14.49 to the sum of the variances of the individual strata. The first approach is more straightforward, and is more easily extendable to more complex situations.

This treatment is equally applicable to an inventory difference as well as a component. If Equation 14.27 is substituted into Equation 14.42 rather than 14.43, the only change in the final result given in Equations 14.46 and 14.47 is that Equation 14.46 becomes

$$\begin{aligned}\mu &= \sum_i c_i\mu_i \\ \mu_{(r)} &= \sum_{i(r)} c_i\mu_i \\ \mu_{j(r)} &= \sum_{i(r)} c_i\mu_{ij(r)}\end{aligned}\quad (14.50)$$

Example 14.5 During two successive 6-month material balance periods, 64 cylinders of fuel material containing 2 metric tons each of UF_6 were added to the process at a low-enriched fuel fabrication facility. For this simplified example, consider a combined measurement process applied to each cylinder that is recalibrated after every eighth cylinder is measured. As in Example 14.4, the constant effects associated with individual measurement and calibration will be combined into a single effect. Assume a relative standard deviation of $\delta_1 = 0.5\%$ for the individual measurements, $\delta_2 = 0.1\%$ for the individual calibrations, and $\theta_1' = \sqrt{\theta_1^2 + \theta_2^2} = 0.05\%$ for the single long-term residual effect. Then for the combined input over the two periods, the estimated variance in the total can be computed using Equation 14.34. Because the total input of 128 metric tons is

assumed to consist of 64 equal item amounts of 2 metric tons each and 8 equal calibration batches of 16 metric tons each, $a_1 = 64$, $a_2 = 8$, and $b_1 = b_2 = 1$. The result is

$$\begin{aligned}\text{Var}(x) &= (128)^2 [(0.0005)^2 + (0.005)^2/64 + (0.001)^2/8] \\ &= 0.012544 (\text{tons})^2\end{aligned}$$

so that the standard deviation of the estimate of the combined input is 112 kg.

If 34 cylinders were processed during the first material balance period and 30 during the second, then the variances for the inputs x_1 and x_2 in the first and second periods, respectively, can be computed using Equation 14.32. Because the final calibration batch of the first period consists of two items and the first calibration batch of the second period consists of six items,

$$\begin{aligned}\text{Var}(x_1) &= (68)^2 (0.0005)^2 \\ &\quad + [34(2)^2] (0.005)^2 \\ &\quad + [4(16)^2 + (4)^2] (0.001)^2 \\ &= 0.005596 (\text{tons})^2 [5596 (\text{kg})^2]\end{aligned}$$

and

$$\begin{aligned}\text{Var}(x_2) &= (60)^2 (0.0005)^2 \\ &\quad + [30(2)^2] (0.005)^2 \\ &\quad + [(12)^2 + 3(16)^2] (0.001)^2 \\ &= 0.004812 (\text{tons})^2 [4812 (\text{kg})^2]\end{aligned}$$

The increase $12,544 - (5596 + 4812) = 2136 (\text{kg})^2$ between $\text{Var}(x)$ and $\text{Var}(x_1) + \text{Var}(x_2)$ is the sum of twice the covariance $(4000)(12,000)(0.001)^2 = 48 (\text{kg})^2$ due to the common calibration effect on the last two items of the first period and the first six items of the second period, and twice the covariance $(68,000)(60,000)(0.0005)^2 = 1020 (\text{kg})^2$ due to the assumption of a common long-term constant effect during the two periods.

Example 14.6 This example will consider the entire simplified material balance for a fuel fabrication facility introduced in Example 14.3.

The other six strata, in addition to the input stratum of UO_2 powder considered there, are as follows:

| Stratum | Description |
|---------|---|
| 2 | An output stratum consisting of containers of sintered UO_2 pellets. |
| 3 | An output waste stream consisting of containers of solid waste measured by NDA. |
| 4,6 | Beginning and ending inventory strata consisting only of containers of dirty scrap. |
| 5,7 | Beginning and ending inventory strata consisting of containers of grinder sludge. |

After eliminating any identical item amounts occurring in two different strata that enter the inventory difference equation with opposite signs, the contents of the strata were as follows:

| Stratum | Total amount
(kg U) | Description |
|---------|------------------------|---|
| 1 | 240,000 | 80 batches of 150 containers of UO_2 powder, each container containing 20 kg of uranium. |
| 2 | 238,800 | 47,760 trays of pellets, each tray containing 5 kg of uranium. |
| 3 | 1,200 | 2,770 containers of solid waste, each containing an average of about 0.433 kg of uranium. |
| 4,6 | 7,200 | 6 batches of 300 containers of dirty scrap, each container containing 4 kg of uranium. |
| 5,7 | 4,000 | 4 batches of 200 containers of grinder sludge, each container containing 5 kg of uranium. |

As described in Example 14.3, the bulk amount in Stratum 1 is determined by weighing each container, but the concentration factor applied to all containers in a batch is based on the analysis of five samples taken at random from that batch. In Stratum 2, the uranium content of all trays is based on the individual weight of each tray of pellets times a "pellet factor" determined from the random selection and analysis of 240 samples of

5 pellets each. In Stratum 3, each solid waste container is individually analyzed by NDA. In Strata 4 through 7, each container of dirty scrap and grinder sludge is weighed separately, but these containers are arbitrarily batched into 6 batches of 300 and 4 batches of 200 containers, respectively. Ten containers of dirty scrap and 12 containers of grinder sludge are selected from each batch for sampling and analysis to determine the concentration factor applied to all items in the batch.

The error sources and their descriptions are as follows:

| Source | Description |
|--------|--|
| 1 | Weighing of containers in Stratum 1;
$\theta_1 = 0.000439, \delta_1 = 0.000658.$ |
| 2 | Sampling and analysis in Stratum 1;
$\theta_2 = 0.000571, \delta_2 = 0.000685.$ |
| 3 | Weighing of pellet trays;
$\theta_3 = 0.000175, \delta_3 = 0.000877.$ |
| 4 | Sampling and analysis of pellets;
$\theta_4 = 0.000341, \delta_4 = 0.000568.$ |
| 5 | NDA measurement of total amount of uranium in a container of solid waste;
$\theta_5 = 0.0462, \delta_5 = 0.0577.$ |
| 6 | Weighing of dirty scrap or grinder sludge;
$\theta_6 = 0.00167, \delta_6 = 0.00250.$ |
| 7 | Sampling and analysis of dirty scrap;
$\theta_7 = 0.00896, \delta_7 = 0.03284.$ |
| 8 | Sampling and analysis of grinder sludge;
$\theta_8 = 0.01000, \delta_8 = 0.04998.$ |

As discussed in Section 13.3.2.2, the specific values of the variance components are computed primarily from data obtained from a control program designed to establish and maintain the quality of the measurements. Chapter 15 will describe the common techniques used for calibration and standardization of measurement methods and the estimation of error variances. In this specific example, the variance components for the weighing operations (Sources 1, 3, and 6) are derived from quality control checks on the scales in routine use, which are carried out at least once per operating shift, and periodic recalibrations by maintenance personnel. The variance components for the sampling and analysis of powder and pellets (Sources 2

and 4) are based on data from the required replicate analysis of standards by laboratory personnel, and from the analysis of the variability of the replicate samples taken to establish powder and pellet factors. The larger contributions of the sampling variances in Strata 4 through 7 (Sources 7 and 8) are established from periodic experiments designed to produce separate estimates of sampling and analytical variance components. The errors of NDA measurement (Source 5) are derived from data from continuing calibration checks against working standards and from experiments designed to estimate the variance due to the distribution of the material in the storage containers. This latter source of error is not obtained through routine checks of a working standard and is the major source of uncertainty in NDA measurements of this type.

Now consider the contribution to $\text{Var}(x)$, as defined by Equation 14.47, of each measurement error source. For $r = 1$, $a_1 = 12,000$, $b_1 = 1$, $N_1 = 1$, $\mu_{(1)} = 240,000$, and $\mu_{j(1)} = \mu_{1k} = 20$ (which is the weight of each individual item in Stratum 1). Thus, the contribution from Source 1 is

$$\begin{aligned}\text{Var}_1(x) &= (240,000)^2(0.000439)^2 + \sum_{k=1}^{12,000} (20)^2(0.000658)^2 \\ &= (240,000)^2(0.000439)^2 + 12,000(20)^2(0.000658)^2 \\ &= 11,100.73 + 2.08 \\ &= 11,102.81 \text{ (kg)}^2\end{aligned}$$

For $r = 2$ (Sources 2 and 3, Example 14.3), $a_2 = 80$, $b_2 = 5$, $N_2 = 150$, $\mu_{(2)} = 240,000$, and $\mu_{j(2)} = \sum_{k=1}^{150} \mu_{1k} = 150(20) = 3000$. The contribution to the variance from this source is

$$\begin{aligned}\text{Var}_2(x) &= (240,000)^2(0.000571)^2 + 80(3000)^2(0.000685)^2/5 \\ &= 18,779.96 + 67.57 \\ &= 18,847.53 \text{ (kg)}^2\end{aligned}$$

Note that the sum of these two contributions, except for a slight variation in the last decimal because of rounding error, is exactly the result of Example 14.3.

For $r = 3$, $a_3 = 47,760$, $b_3 = 1$, and $N_3 = 1$, because each of the 47,760 trays is weighed individually. Hence $\mu_{(3)} = 238,800$, the amount

in the single stratum affected by this source, and $\mu_{j(3)} = \mu_{2k} = 5$, the weight of each individual item in Stratum 2. The contribution from Source 3 is therefore

$$\begin{aligned}\text{Var}_3(x) &= (238,800)^2(0.000175)^2 + 47,760(5)^2(0.000877)^2 \\ &= 1746.40 + 0.92 \\ &= 1747.32 \text{ (kg)}^2\end{aligned}$$

Source 4 again applies only to Stratum 2. There is a single concentration factor corresponding to the one batch of $N_4 = 47,760$ items, and $b_4 = 240$ is the number of replicate samples and analyses used to establish the constant concentration factor applied to all items in the batch. Because the stratum and the single batch are identical, $\mu_{(4)} = \mu_{1(4)} = 238,800$. Thus, the contribution from Source 4 is

$$\begin{aligned}\text{Var}_4(x) &= (238,800)^2[(0.000341)^2 + (0.000568)^2/240] \\ &= 6707.63 \text{ (kg)}^2\end{aligned}$$

Source 5 is a single NDA measurement error applying to each of the $a_5 = 2770$ batches of one item each in Stratum 3; $\mu_{(5)} = 1200$ and $\mu_{j(5)} = \mu_{3k} = 0.433$ so that the contribution from Source 5 is

$$\begin{aligned}\text{Var}_5(x) &= (1200)^2(0.0462)^2 + 2770(0.433)^2(0.0577)^2 \\ &= 3073.59 + 1.73 \\ &= 3075.32 \text{ (kg)}^2\end{aligned}$$

Source 6 affects Strata 4, 5, 6, and 7. Strata 4 and 6 each contain 1800 containers weighed individually. Strata 5 and 7 each contain 800 containers weighed individually. Thus $a_6 = 5200$, $b_6 = 1$, and $N_6 = 1$. Because the beginning inventory strata enter the inventory difference with positive signs, $c_4 = c_5 = +1$ and the ending inventory strata are subtracted, so that $c_6 = c_7 = -1$. Hence,

$$\begin{aligned}\mu_{j(6)} &= c_4\mu_{4k} = +4 \quad \text{for } j(6) = 1 \text{ to } 1800 \\ \mu_{j(6)} &= c_5\mu_{5k} = +5 \quad \text{for } j(6) = 1801 \text{ to } 2600 \\ \mu_{j(6)} &= c_6\mu_{6k} = -4 \quad \text{for } j(6) = 2601 \text{ to } 4400 \\ \mu_{j(6)} &= c_7\mu_{7k} = -5 \quad \text{for } j(6) = 4401 \text{ to } 5200\end{aligned}$$

Thus,

$$\begin{aligned}\mu_{(6)} &= \sum_{i=4}^7 c_i \mu_i \\ &= 7200 + 4000 - 7200 - 4000 \\ &= 0\end{aligned}$$

$$\begin{aligned}\sum_{j(6)} \mu_{j(6)}^2 &= 1800(+4)^2 + 800(+5)^2 \\ &\quad + 1800(-4)^2 + 800(-5)^2 \\ &= 97,600\end{aligned}$$

and

$$\begin{aligned}\text{Var}_6(x) &= (0)^2(0.00167)^2 + 97,600(0.00250)^2 \\ &= 0.61 \text{ (kg)}^2\end{aligned}$$

The same constant relative error effect in equal beginning and ending inventories contributes nothing to the variance of the inventory difference, regardless of the magnitude of the effect.

Source 7 involves Strata 4 and 6. The total number of batches is $a_7 = 12$ (6 in Stratum 4 and 6 in Stratum 6), $b_7 = 10$, and $N_7 = 300$; $\mu_{j(7)} = (+4)(300) = +1200$ for $j(7) = 1$ to 6, and $\mu_{j(7)} = (-4)(300) = -1200$ for $j(7) = 7$ to 12. Thus,

$$\begin{aligned}\mu_{(7)} &= c_4 \mu_4 + c_6 \mu_6 \\ &= +7200 - 7200 \\ &= 0\end{aligned}$$

and

$$\begin{aligned}\sum_{j(7)} \mu_{j(7)}^2 &= 6(1200)^2 + 6(-1200)^2 \\ &= 17,280,000\end{aligned}$$

so that

$$\begin{aligned}\text{Var}_7(x) &= (0)^2(0.00896)^2 + (17,280,000)(0.03284)^2/10 \\ &= 1863.59 \text{ (kg)}^2\end{aligned}$$

Similarly, Source 8 involves Strata 5 and 7, where $a_8 = 8$, $b_8 = 12$, and $N_8 = 200$; $\mu_{j(8)} = (+5)(200) = +1000$ for $j(8) = 1$ to 4,

and $\mu_{j(8)} = (-5)(200) = -1000$ for $j(8) = 5$ to 8 ; $\mu_{(8)} = c_5\mu_5 + c_7\mu_7 = 4000 - 4000 = 0$, and

$$\begin{aligned}\sum_{j(8)}\mu_{j(8)}^2 &= 4(1000)^2 + 4(-1000)^2 \\ &= 8,000,000\end{aligned}$$

Thus,

$$\begin{aligned}\text{Var}_8(x) &= (0)^2(0.01000)^2 + (8,000,000)(0.04998)^2/12 \\ &= 1665.33 \text{ (kg)}^2\end{aligned}$$

Combining the contributions from the eight sources, the variance in the inventory difference is

$$\begin{aligned}\text{Var}(x) &= \sum_{r=1}^8 \text{Var}_r(x) \\ &= 45,010.14 \text{ (kg)}^2\end{aligned}$$

which differs only because of rounding error from the value of 45,010.01 kg^2 obtained in the *Safeguards Technical Manual* (IAEA, 1982; Vol. 3, Part F). The standard deviation of the estimate of the inventory difference is the square root of this estimated variance, or 212 kg.

Note that the existence of single random error effects, which are constant throughout the period in the major process streams, dominates the contributions of measurement error to the variance of the inventory difference in this example. Periodic calibration or standardization of the measurement methods during the period could reduce the variance appreciably.

REFERENCES

- ANSI, 1982, *Statistical Terminology and Notation for Nuclear Materials Management*, ANSI N15.5-1982 (draft), American National Standards Institute, New York.
- Argentis, F., T. Casilli, and M. Franklin, 1979, Nuclear Material Statistical Accountancy System, *European Applied Research Reports*, Harwood Academic Publishers, Chur, Switzerland, 1(5): 1085-1180.
- Byers, K. R., J. W. Johnston, C. A. Bennett, R. J. Brouns, M. F. Mullen, and F. P. Roberts, 1983, *Assessment of Alternatives to Correct Inventory Difference Statistical Treatment Deficiencies*, NUREG/CR-3256, U. S. Nuclear Regulatory Commission, Washington, D. C., available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- DeBieve, P., 1984, 1983 Target Values for Uncertainty Components in Fissile Element and Isotope Assay, *ESARDA Bulletin*, No. 6, April 1984, Joint Research Centre, I-21020, Ispra, Italy.
- Franklin, M., 1984, The Sampling Distribution of the D Statistic, *J. Inst. Nucl. Mater. Manage.*, Proceedings Issue, 25th Annual Meeting, XIII: 135-142.

- Hakkila, E. A., C. C. Thomas, Jr., D. B. Smith, C. A. Ostenak, and R. G. Gutmacher, 1980, Conventional Measurement Techniques, *Materials Management in an Internationally Safeguarded Fuels Reprocessing Plant*, LA-8042, Vol. III, App. K, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- IAEA, 1982, Statistical Concepts and Techniques, *IAEA Safeguards Technical Manual*, IAEA-TECDOC-261, Vol. 3, Part F, International Atomic Energy Agency, Vienna.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Lumb, R. F., M. Messinger, and F. H. Tingey, 1983, Augmented Automated Material Accounting Statistics System (AMASS), *J. Inst. Nucl. Mater. Manage.*, Proceedings Issue, 24th Annual Meeting, XII: 221-226.
- Messinger, M., R. F. Lumb, and F. H. Tingey, 1981, Automated Material Accounting Statistics System (AMASS), *J. Inst. Nucl. Mater. Manage.*, Proceedings Issue, 22nd Annual Meeting, X: 125-131.
- Reilly, T. D., and M. L. Evans, 1977, *Measurement Reliability for Nuclear Material Assay*, LA-6574, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Rogers, D. R. (Ed.), 1983, *Handbook of Nuclear Safeguards Measurement Methods*, NUREG/CR-2078, U. S. Nuclear Regulatory Commission, Washington, D. C., available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Stirpe, D., A. S. Goldman, and R. G. Gutmacher, 1983, *MC&A Reform Amendment: Example System Beta Variance Estimates and Performance Evaluation*, NUREG/CR-3221 (LA-9708-MS), U. S. Nuclear Regulatory Commission, Washington, D. C., available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Tingey, F. H., and R. F. Lumb, 1984, *Validation of the Generic Measurement Structure Employed in AMASS*, NUSAC Report 915, NUSAC, Inc., Reston, Virginia.

APPENDIX 14A

An Example of Modeling Measurement Variability

14A.1 INTRODUCTION

Throughout this chapter the generic nature of the treatment of sources of measurement variability is emphasized. The methods given to describe and compute estimates of the total variability due to the identified error sources are completely general. The extent to which it is necessary to identify and take into account individual sources depends on the specific application.

The following example describes the measurement process for determining the input to a scrap recovery operation and the specific sources of error present. It also describes the modeling process for the variability due to one part of the measurement process, the net weight determination.

14A.2 MEASUREMENT SYSTEM DESCRIPTION

An input accountability measurement system for a scrap recovery operation is shown in Figure 14A.1. Assume that the material to be processed is dumped into a dissolver tray and dissolved with nitric acid, passed through a filter and transferred to a holding tank by pulling a vacuum on the holding tank. When the holding tank has accumulated about 35 kg of solution, the solution is transferred over to the weigh tank. The weigh tank is suspended from a load cell with a capacity of 98 kg. When the transfer is completed, a gross weight reading is taken by observing the digital readout from the load cell. The digital readout is set to read to the nearest 20 g. After the gross reading has been taken, the solution is mixed by connecting an air hose to a valve at the bottom of the tank and air sparging for 30 min to assure chemical homogeneity of the solution. Next, the two sampling ports A and B of Figure 14A.1 are rinsed by opening the ports and drawing off some solution into a bucket. Two samples representing the transfer are then taken by nearly filling a 4 oz (100 ml) plastic bottle from each port. After the sampling process is complete, the solution is transferred to the feed tank and the tare reading of the weigh tank taken.

Some other pertinent characteristics of the weighing process are:

1. About 35 kg of solution is contained in each transfer.
2. The uranium concentration is usually less than 10 g/kg so that less than 350 g of uranium is contained in each transfer.

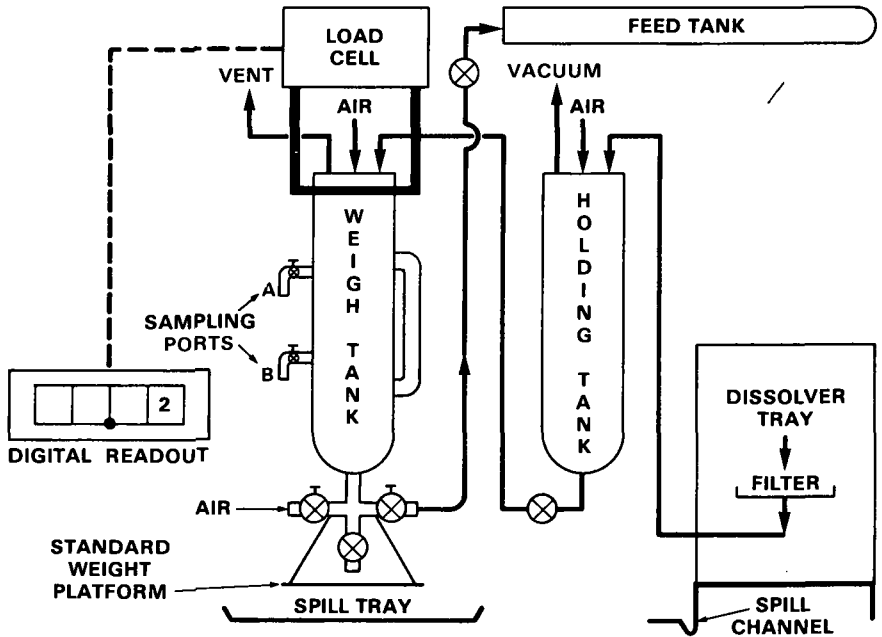


Figure 14A.1 Schematic of input accountability system.

3. Each transfer takes about 40 min (from holding tank, through weigh tank, to feed tank).

4. The digital readout may fluctuate by as much as ± 60 g during the gross and tare readings for three of the scales and by ± 20 g for the fourth scale.

5. The tare weight is preset electronically at a nominal value of 3 to 6 kg. The actual tare of the empty tank is about 55 kg.

6. There is a 30-min interval between the taking of the gross and tare readings.

Measurement control practice for the weight determination is to check each scale daily with a 30 kg standard. The tare reading is set so that the empty tank registers 3 to 6 kg on the readout. This reading is recorded, the 30 kg test weight is placed on a platform hung from the valve tree at the bottom of the tank, and the gross weight recorded. The data collected from these check weighings are used to estimate error variances at the end of each accountability period.

The operating practice for uranium concentration determinations is to composite all samples from the A-port and all samples from the B-port, and draw an aliquot from each composite for analysis. The two results are averaged and the result is the grams of uranium per gram of solution

applicable to all the tankfuls run through a specific weigh tank for a specific job. Each job may have several pairs of composited samples depending on how many dissolver-weigh tank systems were used in processing a particular batch of scrap.

The measurement control program for the uranium analysis contains two procedures from which variance components associated with the element analysis can be estimated. The first is used to control bias and estimate the contribution of the calibration process. Working standard solutions, prepared from NBS standards, are run about 40 times during an accountability period. As many as 7 different standards, ranging from 1.9 to 11.7 g U/kg, may be used depending on the expected concentrations of the process solutions to be measured on the day that the standard is run. The second measurement control procedure involves replication in order to estimate the measurement error variances for production measurements. The usual practice is to select eight pairs of composite samples for each weigh tank and determine the concentration of uranium in an aliquot from each of the A and B composites of samples from transfers for the job run through the weigh tank. The measurement results from these duplicate, composited samples for each job are used to estimate the variability due to the combination of sampling and analytical random errors. A second aliquot is taken from either the A or B sample composite container, and the difference between the two results from a single composite is used to estimate the random error for analysis only.

This description of the weighing and analysis operations identifies the potential sources of error in determining the amount of uranium input to the facility. There are six strata of concern corresponding to the following types of scrap:

1. Filter sludge
2. Spilled or leaked solution
3. Process holdup in tanks and piping
4. Sample port rinse
5. Samples of solution
6. Recycle added to process through the accountability tank

This example concentrates on Stratum 6. In practice, a similar modeling process would be required for each stratum.

There are numerous sources of variability for the weigh tank and concentration measurements. Load cell variations due to temperature, source voltage fluctuations, and observer averaging to the nearest 20 g can be sources of individual error effects in each measurement. Potential sources of constant error effects include load cell hysteresis (delay in returning to calibrated state after a load is applied), the age of the electronics in a corrosive atmosphere, and operator reading bias (a "company man" may read low to protect against shipper-receiver differences). Sampling errors due to

lack of solution homogeneity can contribute several types of variability. For example, rinsing sampling ports with the solution in the current tankful is designed to eliminate serially correlated effects due to cross-contamination with solution from previous tankfuls.

These sources of error do not all have to be explicitly modeled. However, their existence should be recognized so that sources included in the modeled errors are understood and data collection experiments to estimate the associated variability can be properly designed. The consideration of models for the measurement control processes anticipates the techniques described in Chapter 15 for estimating and controlling measurement variability.

14A.3 MODELING THE WEIGHING OPERATION

For each stratum, the net weight and uranium concentration are the measurement results y_1 and y_2 necessary to determine the amount x of uranium contained in each tankful. The division into strata corresponding to various types of scrap is required because of expected differences in the variability associated with the measurement of uranium concentration. The following sections deal with the modeling of the measurement variability in the net weight; this model will apply equally to all strata. Because only the measurement of the bulk amount is considered, the "1" subscript on the net weight measurement will be dropped. Finally, the development will apply to all four weigh tanks, since the operating and control procedures, and thus the error sources to be modeled, are the same in all systems. However, the estimated variance components, and hence the estimated variability in the net weight, may differ from system to system.

14A.3.1 Constructing the Error Model

Let y_k be the observed net weight for tankful (item) k processed in this material balance period. As noted in the description of the measurement system in the previous section, the net weight transferred is the difference between a gross weight and a tare weight, so that

$$y_k = Y_{Gk} - Y_{Tk} \quad (14A.1)$$

The observed gross and tare weights can be written as

$$\begin{aligned} Y_{Gk} &= \mu_{Gk} + \Delta_{Gk} \\ Y_{Tk} &= \mu_{Tk} + \Delta_{Tk} \end{aligned} \quad (14A.2)$$

where μ_{Gk} and μ_{Tk} are the true gross and tare weights of the k^{th} tankful and Δ_{Gk} and Δ_{Tk} are the associated weighing errors. Both the errors Δ_{Gk}

and Δ_{Tk} may be modeled as the sum of a constant error effect associated with a defined subset of N_j tankfuls (items) and a residual error effect associated with the particular item, so that

$$\begin{aligned}\Delta_{Gk} &= \beta_{Gj} + \epsilon_{Gk} \\ \Delta_{Tk} &= \beta_{Tj} + \epsilon_{Tk}\end{aligned}\quad (14A.3)$$

where $j = 1, \dots, a$ is the number of subsets weighed in a particular system during the material balance period. This introduction of a constant error effect associated with a defined subset is meaningful only if the errors are not completely random but change systematically from time period to time period or with other characteristics of the weighing procedure. The particular classification into subsets associated with process jobs (batches) is not particularly useful for modeling and estimating variability due to weighing. Other classifications of the tankfuls (items) into subsets (batches) associated with the measurement control program may be more meaningful.

Since the gross and tare measurements on production tankfuls are not replicated, error estimation must be based on data from current measurement control programs or, when necessary, historical data from prior control programs or interlaboratory exchanges. In this specific example, the measurement control data for weighing consist of results of check weighings, using a working standard weight of 30 kg, each day the weigh tank is used, which may not be consecutive days. The procedure involves setting and recording a pre-tare reading (including the platform and chains for suspending the test weight), placing a 30 kg working standard weight on the platform and recording the scale reading, removing the 30 kg weight and recording the first post-tare reading, and recording the second post-tare reading after waiting 5 min. This provides replication over time on a single standard item.

The model for the gross weights obtained from the check weighing of the 30 kg weight on day l is

$$Z_{G\ell} = \nu + \nu_{T\ell} + \beta_{G\ell} + \epsilon_{G\ell} \quad (14A.4)$$

where

- ν = true weight of the 30 kg working standard
- $\nu_{T\ell}$ = true tare weight on day l
- $\beta_{G\ell}$ = constant error in measuring gross weight of 30 kg working standard on day l
- $\epsilon_{G\ell}$ = random error in measuring gross weight of 30 kg working standard on day l

In the single measurement of the gross weight of the standard on day ℓ , $\beta_{G\ell}$ and $\epsilon_{G\ell}$ are not distinguishable. The distinction is that $\beta_{G\ell}$ is assumed to remain fixed for all gross weights determined on day ℓ , while $\epsilon_{G\ell}$ is different for each replicate measurement of a gross weight.

Another set of measurement control data is generated periodically when the 30 kg working standard is compared with NBS certified class-3 weights on a pan balance. The model for the m^{th} calibration weighing of the 30 kg working standard is

$$z_m = \nu + \beta_c + \epsilon_m \quad (14A.5)$$

where

- ν = true weight of the 30 kg working standard
- β_c = constant effect associated with calibration of working standard
- ϵ_m = random effect associated with calibration of working standard

Finally, two types of error are reported by NBS (Ku, 1969, pp. 43, 336) for the value assigned to each class-S standard, so that

$$z_s = \nu_s + \epsilon_s + |\Delta_s| \quad (14A.6)$$

where

- z_s = value assigned to a class-S standard
- ν_s = true value of the standard
- ϵ_s = random error in NBS determination of z_s
- $|\Delta_s|$ = "reasonable bound" to the systematic error

The final set of data generated by the control program is comprised of the duplicate tare measurements taken after each measurement of the gross weight of the 30 kg working standard. The pre-tare data are not used in the data analysis. These duplicate measurements not only allow the direct estimation of random weighing error, but also enable a check for any hysteresis effect. An appropriate model is

$$z_{TR1} = \nu_{TR} + \beta_{TR} + \delta_T + \epsilon_{TR1}$$

$$z_{TR2} = \nu_{TR} + \beta_{TR} + \epsilon_{TR2}$$

where

- ν_{TR} = true tare weight on day ℓ (Equation 14A.4)
- β_{TR} = constant error in measuring tare weight on day ℓ
- $\epsilon_{TR1}, \epsilon_{TR2}$ = random errors in measuring tare weight on day ℓ
- δ_T = possible constant difference between first and second tare measurements

Two specific sets of results derived from these duplicate measurements are of interest. The differences

$$\begin{aligned}\Delta z_{T\bar{L}} &= z_{T\bar{L}1} - z_{T\bar{L}2} \\ &= \delta_T + (\epsilon_{T\bar{L}1} - \epsilon_{T\bar{L}2})\end{aligned}\quad (14A.7)$$

provide the basis for a classical paired difference test to determine if a hysteresis effect δ_T exists (Sections 4.4 and 15.3.2). If the hypothesis $\delta_T = 0$ (no hysteresis effect) is not rejected, the tare weight to be associated with the gross weight $z_{G\bar{L}}$ (Equation 14A.4) will be the average

$$\begin{aligned}z_{T\bar{L}} &= \frac{1}{2} (z_{T\bar{L}1} + z_{T\bar{L}2}) \\ &= \nu_{T\bar{L}} + \beta_{T\bar{L}} + (\epsilon_{T\bar{L}1} + \epsilon_{T\bar{L}2})/2\end{aligned}\quad (14A.8)$$

If the hypothesis $\delta_T = 0$ is rejected, the second post-tare reading is assumed to have no hysteresis effect and

$$z_{T\bar{L}} = z_{T\bar{L}2} = \nu_{T\bar{L}} + \beta_{T\bar{L}} + \epsilon_{T\bar{L}2}$$

14A.3.2 Combining Errors and Taking Expectations

The quantity of interest to the material accounting process is the net amount of material transferred. Under the assumption that all error effects associated with net weight determination are independent of those associated with measurement of concentration, it was shown in Section 14.2.2 that the relative error effects—and hence, the relative variance components—due to net weight measurements are an additive part of the total relative error effect. It follows from Equation 14.18 that the errors are determined relative to the particular measurement involved, even though they ultimately contribute to the relative error in the amount transferred. The net effect is that in the present example, the impact of the weighing operation on the variability in the net amount transferred can be determined by estimating relative variance components with respect to amounts weighed and applying them to amounts transferred. The problem is further simplified in this particular example by the fact that net amounts transferred are relatively constant.

From Equations 14A.1, 14A.2, and 14A.3, the net amount transferred for tankful (item) k is

$$\begin{aligned}y_k &= y_{Gk} - y_{Tk} \\ &= \mu_k + \beta_j + \epsilon_k\end{aligned}\quad (14A.9)$$

where

$$\mu_k = \mu_{Gk} - \mu_{Tk} = \text{true net weight transferred}$$

$$\beta_j = \beta_{Gj} - \beta_{Tj} = \text{constant error effect associated with subset } j \text{ of the items measured}$$

$$\epsilon_k = \epsilon_{Gk} - \epsilon_{Tk} = \text{individual error effect in net weight of an item due to individual error effect in gross and tare weight measurements}$$

Two approaches are possible to express model 14A.9 for a net weight determination in terms of relative error effects. Both approaches are applicable to either the constant error effects β_j or the individual error effects ϵ_k , so the constant error effects will be used for illustration. The first approach assumes a relative error effect β'_j to be an observed value of a single random variable which is applicable to the weighing of any absolute amount of material associated with the subset j . Thus

$$\beta_{Gj} = \mu_{Gk} \beta'_j$$

$$\beta_{Tk} = \mu_{Tk} \beta'_j$$

and

$$\beta_j = \mu_{Gk} \beta'_j - \mu_{Tk} \beta'_j$$

$$= (\mu_{Gk} - \mu_{Tk}) \beta'_j$$

$$= \mu_k \beta'_j$$

(14A.10)

The second approach assumes that β_j is the difference of two absolute errors β_{Gj} and β_{Tj} , each of which is an observed value on a random variable whose distribution may depend on the weight of the material being weighed. For gross and tare weights that are approximately constant, the distributions of β_{Gj} and β_{Tj} (and hence the difference β_j) will be approximately the same. Because under these conditions the true net weight μ_k is also approximately constant, the relative error $\beta'_j = \beta_j / \mu_k$ will again be an observed value of a random variable that is approximately identically distributed for all weighings on a particular system. The first approach justifies the transformation $\beta_j = \mu_k \beta'_j$ from absolute to relative error effect on the basis that the constant error effect is truly relative. The second approach justifies the same transformation on the basis that absolute and relative errors will be proportional if the gross, tare, and net amounts being weighed are equal for all measurements. This argument is a specific example of one of the bases for the use of relative errors discussed in Section 14.2.1.

By a similar argument, it can be assumed that $\epsilon = \mu_k \eta_k$ for the individual errors, and hence in terms of relative error effects Equation 14A.9 becomes

$$\begin{aligned} y_k &= \mu_k + \mu_k \beta'_j + \mu_k \eta_k \\ &= \mu_k (1 + \beta'_j + \eta_k) \end{aligned} \quad (14A.11)$$

The quantities

$$z_R = z_{GR} - z_{TR}$$

are also of interest in using the data obtained by weighing the 30 kg working standard. From Equations 14A.4 and 14A.8, if no hysteresis effect is assumed,

$$z_R = \nu + (\beta_{GR} - \beta_{TR}) + [\epsilon_{GR} - (\epsilon_{TR1} + \epsilon_{TR2})/2] \quad (14A.12)$$

Similar arguments also apply to the effects

$$\begin{aligned} \beta_R &= \beta_{GR} - \beta_{TR} \\ \epsilon_R &= \epsilon_{GR} - (\epsilon_{TR1} + \epsilon_{TR2})/2 \end{aligned} \quad (14A.13)$$

so that Equation 14A.12 can be written in terms of the relative errors $\beta'_R = \beta_R/\nu$ and $\eta_R = \epsilon_R/\nu$ as

$$z_R = \nu(1 + \beta'_R + \eta_R) \quad (14A.14)$$

If a hysteresis effect is present, and only the second tare reading is used, the only change in this formulation is to replace $(\epsilon_{TR1} + \epsilon_{TR2})/2$ by ϵ_{TR2} in Equation 14A.13. In this example the gross, tare, and net weights are essentially unchanged for each determination of net weight, so the two approaches to relative error definition produce completely equivalent models for the error effects.

The precise pan-balance calibration of the working standard against NBS standard weights modeled by Equation 14A.5 does not involve gross and tare weight differences. The error β_c and the errors ϵ_m can be expressed as relative errors by dividing them by the true value ν of the working standard, so that

$$z_m = \nu(1 + \beta'_c + \eta_m) \quad (14A.15)$$

where $\beta'_c = \beta_c/\nu$ and $\eta_m = \epsilon_m/\nu$.

The production results modeled by Equation 14A.11, the daily system calibration modeled by Equation 14A.14, and the occasional calibrations of the working standard modeled by Equation 14A.15 constitute the available data from production measurement and the measurement control program during a given material balance period. If the subset j of the production measurements is chosen to represent one day's weighings, then the constant error effects β'_c of Equation 14A.15 (β_j of Equation 14A.9) are equivalent to the constant error effects β'_ℓ in the daily calibration check modeled in Equation 14A.14. With this identification it is possible, following Section 14.2.1, to define three sources of variability in the net weight determination:

1. Variability associated with repeated individual determination of a gross or tare weight on a production sample or standard working standard.
2. Variability in the constant deviation from the standard value of all weighings during a particular day.
3. Variability in the assignment of a standard value to the working standard.

Using the notation of Section 14.2.3, these three sources will be identified by the indices

$$\begin{aligned} j(1) (= k) &= 1, \text{-----}, a_1 = N \\ j(2) (= \ell) &= 1, \text{-----}, a_2 \\ j(3) (= m) &= 1, \text{-----}, a_3 \end{aligned}$$

where N is the number of weighing operations, a_2 is the number of operating days in the material balance period, and a_3 is the number of calibrations of the working standard that were in effect during the material balance period.

There are two ways to use the observed data to correct the weight of material transferred in an individual tankful. Both involve correcting the production data on the basis of differences between the daily check measurements and the assigned value of the working standard. From Equations 14A.12 and 14A.5, the absolute differences are

$$\begin{aligned} \Delta_{\ell m} &= z_\ell - z_m \\ &= \beta_\ell - \beta_c + \epsilon_\ell - \epsilon_m \end{aligned} \quad (14A.16)$$

The first approach is to correct all the n_ℓ weighings on a single day by the individual difference $\Delta_{\ell m}$. It then follows from Equations 14A.9 and 14A.16 that

$$\begin{aligned} \hat{\mu}_k &= y_k - \Delta_{\ell m} \\ &= \mu_k + \beta_j - \beta_\ell + \beta_c + \epsilon_k - \epsilon_\ell + \epsilon_m \end{aligned} \quad (14A.17)$$

The key assumption at this point is that the constant error effect β_ℓ associated with the check measurement on day ℓ is identical to the constant effect β_j in the subset of production measurements made on that day. In that case

$$\hat{\mu}_k = \mu_k + \beta_c + \epsilon_k - \epsilon_\ell + \epsilon_m \quad (14A.18)$$

and the only remaining constant error effect is the possible long-term bias β_c in the assignment of values to the working standard. This bias will include any systematic error in the "reasonable bound" associated with the NBS standard weights. The usual practice is to assume that the quality of the pan-balance procedure and the NBS certification procedure is such that this constant error effect is negligible.

In cases where immediate application of individual corrections is not feasible, an alternate practice frequently used in material accounting is to apply an after-the-fact correction to the total amount of material transferred based on the average difference for the period between the check weighings and the assigned value of the standard. From Equation 14A.16, the average correction over the a_2 days in a material balance period is

$$\begin{aligned} \bar{\Delta}_m &= \sum_{\ell=1}^{a_2} \Delta_{\ell m} / a_2 \\ &= \bar{\beta}_2 - \beta_c + \bar{\epsilon}_2 - \bar{\epsilon}_3 \end{aligned} \quad (14A.19)$$

where

$$\bar{\beta}_2 = \left(\sum_{\ell=1}^{a_2} \beta_\ell \right) / a_2 \quad (14A.20)$$

$$\bar{\epsilon}_2 = \left(\sum_{\ell=1}^{a_2} \epsilon_\ell \right) / a_2 \quad (14A.21)$$

$$\bar{\epsilon}_3 = \left(\sum_{m=1}^{a_1} c_m \epsilon_m \right) / a_2 \quad (14A.22)$$

and c_m is the number of days the m^{th} assigned value of the 30 kg standard was used to determine $\Delta_{\ell m}$, so that

$$\sum_{m=1}^{a_1} c_m = a_2 \quad (14A.23)$$

Because in this case

$$\hat{\mu}_k = y_k - \bar{\Delta}_m$$

it follows from Equations 14A.9 and 14A.19 that

$$\hat{\mu}_k = \mu_k + \beta_j - \bar{\beta}_2 + \beta_c + \epsilon_k - \bar{\epsilon}_2 + \bar{\epsilon}_3 \quad (14A.24)$$

In this case $\bar{\beta}_2$ (the average of the constant daily error effects associated with the check measurements) is not identically equal to the constant daily error effects β_j associated with production measurements, even if the assumption is made that on a given day the two effects are identical. The variability associated with the deviations $\beta_j - \bar{\beta}_2$ of the constant daily error effects from their average is still present in the results. On the other hand, the variability associated with the average errors $\bar{\epsilon}_2$ and $\bar{\epsilon}_3$ will be smaller than that associated with the ϵ_l and ϵ_m of Equation 14A.18.

The analogous development, using relative error models, illustrates the need for assuming either the universal applicability of the relative error distribution to all weights or the equality of the amounts being measured. If Δ_{lm} is determined from Equations 14A.14 and 14A.15, then Equation 14A.16 becomes

$$\Delta_{lm} = \nu(\beta'_l - \beta'_c + \eta_l - \eta_m) \quad (14A.25)$$

To combine this difference with the relative error model for y_k given by Equation 14A.11 to produce the relative error model

$$\hat{\mu}_k = \mu_k(1 + \beta'_j - \beta'_l + \beta'_c + \eta_k - \eta_l + \eta_m) \quad (14A.26)$$

comparable to the absolute model given by Equation 14A.17, requires that either (1) $\mu_k = \nu$ for all k or (2) the relative errors β'_l , β'_c , η_l , and η_m , defined with respect to the standard value ν , apply to all magnitudes of the true contents μ_k of the production items measured. The first of these assumptions is obviously untenable, but it emphasizes the need to have a working standard whose weight is similar to the net weight of an individual operating transfer. In either case the equality $\beta'_j = \beta'_l$ can be assumed, either directly or as a consequence of the assumption $\beta_j = \beta_l$, so that the model reduces to the form

$$\hat{\mu}_k = \mu_k (1 + \beta'_c + \eta_k - \eta_l + \eta_m) \quad (14A.27)$$

analogous to Equation 14A.18. A relative error model

$$\bar{\Delta}_m = \nu(\bar{\beta}'_2 - \beta'_c + \bar{\epsilon}'_2 - \bar{\epsilon}'_3) \quad (14A.28)$$

equivalent to Equation 14A.19 can be obtained following the definitions in Equations 14A.19 through 14A.22 because all the errors involved are relative to the constant true value ν of the working standard. However, combining this model with Equation 14A.11 to obtain

$$\hat{\mu}_k = \mu_k(1 + \beta'_j - \bar{\beta}'_2 + \beta'_c + \eta_k - \bar{\eta}_2 + \bar{\eta}_3) \quad (14A.29)$$

which is the relative error model analogous to the absolute error model given by Equation 14A.24, requires the same assumptions used to develop Equation 14A.26.

14A.3.3 Defining the Variance Components

This section considers the assumed statistical properties of the errors associated with each source of variability present in the models of the previous section and defines the variance components for each source based on the generic structures of Section 14.3.1. Consider as the first source the variability associated with the replication of the weighing operation within a short time period. A frequently justified assumption is that the errors ϵ_{Gk} and ϵ_{Tk} defined in Equation 14A.3 are each identically and independently distributed for all k with expected values β_{1G} and β_{1T} , variances σ_{1G}^2 and σ_{1T}^2 , and zero covariance. Then the differences $\epsilon_k = \epsilon_{Gk} - \epsilon_{Tk}$ are independently and identically distributed with expected value $\beta_1 = \beta_{1G} - \beta_{1T}$ and variance $\sigma_1^2 = \sigma_{1G}^2 + \sigma_{1T}^2$. If it is assumed that the distribution of absolute errors in gross and tare weight determinations due to this source is the same, then $\beta_1 = 0$ and $\sigma_1^2 = 2\sigma_{1G}^2 = 2\sigma_{1T}^2$. If the relative errors $\eta_{Gk} = \epsilon_{Gk}/\mu_{Gk}$ and $\eta_{Tk} = \epsilon_{Tk}/\mu_{Tk}$ are each identically and independently distributed with expected values β'_{1G} and β'_{1T} , variances δ_{1G}^2 and δ_{1T}^2 , and covariance zero, then the relative errors η_k have expected values

$$\beta'_{1k} = (\mu_{Gk}\beta'_{1G} - \mu_{Tk}\beta'_{1T})/\mu_k \quad (14A.30)$$

and variances

$$\delta_{1k}^2 = \mu_{Gk}^2 \delta_{1G}^2 / \mu_k^2 + \mu_{Tk}^2 \delta_{1T}^2 / \mu_k^2 \quad (14A.31)$$

If it is assumed that $\beta'_{1G} = \beta'_{1T}$ and $\delta^2_{1G} = \delta^2_{1T}$, then the expected values

$$\beta'_{1k} = \beta'_{1G} = \beta'_{1T} = \beta'_1 \quad (14A.32)$$

are constant for all k , but the variances

$$\delta^2_{1k} = (\mu^2_{Gk} + \mu^2_{Tk})\delta^2_{1G}/\mu^2_k \quad (14A.33)$$

are still dependent on the ratios $\mu_{Tk}/\mu_k = \lambda_k$, and will equal a common value δ^2_1 only if the ratios are constant. For a constant ratio $\lambda_k = \lambda$, the common variance is

$$\delta^2_1 = (1 + \lambda)^2 \delta^2_{1G} + \lambda^2 \delta^2_{1T} \quad (14A.34)$$

As λ approaches zero, δ^2_1 approaches δ^2_{1G} . The relative variability in net weight will usually be somewhat larger than the combined relative variability in the gross and tare weights.

All these considerations affect the validity of the assumption that the individual relative error effects associated with the weight of the material in the n tankfuls transferred have the same expected value and variance. The key criterion for the validity of this assumption, under a wide variety of assumptions concerning the basic error structure, is the approximate equality of the net weights for the n tankfuls transferred. However, the specific assumptions concerning the error structure may have a significant effect on the techniques used to estimate the error variance components from available experimental or control data.

The only quantities in this example whose variability depends solely on the variability in individual weighings are the absolute differences in tare weights defined by Equation 14A.7. The sample variance of these differences is an estimate of $2\sigma^2_{1T}$. It can be used to estimate δ^2_1 , the relative variance in a single net weight due to this source, only under one of two possible assumptions analogous to those discussed in connection with the development of the relative error model (Equation 14A.26). One possibility is to assume that $\sigma^2_{1G} = \sigma^2_{1T}$. Then $\hat{\sigma}^2_1 = 2\hat{\sigma}^2_{1T}$ and $\hat{\delta}^2_1 = \hat{\sigma}^2_1/\mu^2$, where μ is the average value of the approximately equal net weights μ_k . The second possibility is to assume that the estimated relative error variance $\hat{\delta}^2_{1T} = \hat{\sigma}^2_{1T}/\nu^2_T$, where ν_T is the average of the approximately equal tare weights ν_{Tj} associated with the daily control checks, can be applied to all weighings. Then both δ^2_{1G} and δ^2_{1T} in Equation 14A.34 can be replaced by $\hat{\delta}^2_{1T}$ to obtain the estimate

$$\hat{\delta}^2_1 = [(1 + \lambda)^2 + \lambda^2]\hat{\delta}^2_{1T}$$

where $\lambda = \mu_T/\mu$ is the approximately constant ratio of tare to net weights for production items.

The ratio of the estimate obtained from the second approach to that obtained from the first is $[(1 + \lambda)^2 + \lambda^2]\mu^2/2\nu_T^2$. If $\mu_G = \mu + \mu_T$ is the approximately constant gross weight, and $\mu_T \cong \nu_T$, then this ratio reduces to $(1 + \tau^2)/2$, where $\tau = \mu_G/\mu_T$ is the ratio of gross to tare weight. When the tare weight is small compared to the net transfer, this ratio can be quite large. The two estimates are equal only for $\tau = 1$, which corresponds to $\mu = 0$. This is the only condition under which the gross and tare weighings can have equal absolute and relative errors. The other extreme is the case where μ_T is small compared to μ_G , in which case the second approach gives much larger estimates of the relative error in a net weight. The question is whether the first approach underestimates because of the assumption that the absolute error in a large gross weight is the same as the absolute error in a smaller net weight, or whether the second approach overestimates because the large relative error with respect to a small tare weight is not applicable to the larger gross weight. This is an example of the general problem of extrapolating absolute or relative errors considered in Section 14.2.1. As noted there, the true situation is usually intermediate to the two extremes. The absolute variance of the gross measurements will be higher than that of the tare measurements, but not by the ratio $(\mu_G/\mu_T)^2$.

A similar set of conditions applies to the definition of the expected value β'_2 and the variance δ_2^2 associated with the relative error effects $\beta'_j = \beta_j/\mu_k$ derived from the absolute error effects β_{Gj} and β_{Tj} defined in Equation 14A.3. There is no reason why the distribution of these constant daily error effects should be the same at gross and tare levels on either a relative or absolute basis. Furthermore, the nature of this source is such that there is no way to distinguish between the expected value β'_1 associated with all individual weighings and the expected value β'_2 associated with the constant effect in a day's measurements. In the notation of Section 14.2.3, Equation 14A.9 could have been written

$$y_k = \mu_k + \beta_1 + \epsilon_{j(2)} + \epsilon_{j(1)} \quad (14A.35)$$

emphasizing the hierarchical nature of the error effects. This is also consistent with the fact that only a single long-term average error effect can be estimated from the differences Δ_{lm} defined in Equation 14A.16. If, analogous to Equation 14A.35, Equation 14A.16 is written as

$$\Delta_{lm} = \beta_1 - \beta_3 + \epsilon_{j(2)} + \epsilon_{j(1)} - \epsilon_{j(3)} \quad (14A.36)$$

where $\beta_3 = \beta_c$ and $\epsilon_{j(3)} = \epsilon_m$ refer to the third source of variability due to

calibration of the working standard, it is apparent that this single long-term bias effect depends on the constant error effect present in the calibration of the working standard. The variance estimate obtained from these differences will contain contributions indistinguishable from the variance of the individual error effects from all sources.

The preceding discussion of estimation possibilities also included an implicit assumption that the error effects present in the check measurements on the working standard have the same distribution as the corresponding error effects in production measurements. It is necessary to assume not only that the constant error effects present on a given day are identical, as was done to derive Equation 14A.18 from Equation 14A.17, but also that the variability in the constant error effects β_j and β_R and the individual effects ϵ_k and ϵ_L can be represented by the same variance components. Some implications of the latter assumption are discussed again in Section 15.1.1 in connection with estimation procedures associated with monitoring against standards. This assumption makes it possible to estimate the variability in the total transfer using variance components estimated from data obtained during the check procedure. In particular, the variance σ_2^2 of the daily constant error effects $\epsilon_{j(2)}$ can be estimated by subtracting estimates of σ_1^2 and σ_3^2 from the sample variance of the differences Δ_{km} (see Chapter 15) and σ_3^2 can be estimated from historical variability in the values assigned to the working standard. Under the usual assumptions, both estimates can be converted to estimates of the relative variances δ_2^2 and δ_3^2 by dividing by ν^2 .

As discussed in Section 14.3.1, the most difficult problem is to formulate appropriate assumptions concerning the properties of $\beta_1(\beta'_1)$ and $\beta_3(\beta'_3)$. It is reasonable to expect that, because of the nature of the pan balance calculation and the care taken by NBS establishing standard values, β_3 will be negligible. The only known variance component is the estimate of the variance due to random error provided in the NBS certification. Some idea of the variability in β_1 can be obtained from historical data on control checks for past material balance periods, but such data build up slowly and the pertinence of past data can be questioned. Another source of information may be differences in the long-term average bias for different weighing systems of the same type. Such period-system comparisons are similar to the interlaboratory experiments discussed in Chapter 15.

14A.3.4 Calculating the Total Variability

Using the procedures defined in Sections 14.3.2 and 14.3.3, the variance in the total weight of material transferred can be determined from the error models derived in the preceding section for the two different approaches to correction. The variance when no correction is applied ($\hat{\mu}_k = y_k$) is derived first from the model for y_k in Equations 14A.9 and

14A.11. The first approach to correction considered is an average correction, for which the appropriate models are Equations 14A.24 and 14A.29. The second approach considered is the application of a daily correction factor to individual weights, for which the appropriate models are given by Equations 14A.18 and 14A.27.

Following the conventions established in Section 14.3, define

$$\mu = \sum_{k=1}^N \mu_k$$

as the true value of the total weight of material transferred during the material balance period, $\mu_{j(2)} = \mu_j$ as the total weight of the $N_{j(2)} = N_j$ transfers on each of the a_2 days of the material balance period, and $\mu_{j(3)} = \mu_m$ as the total weight of material transferred during each of the a_3 periods corresponding to a particular assigned value of the working standard. Then if model 14A.11 is written as

$$y_k = \mu_k (1 + \beta'_1 + \eta_{j(2)} + \eta_{j(1)})$$

in analogy to Equation 14A.35, it follows from Equation 14.33 that the variability in the total

$$y = \sum_{k=1}^N y_k$$

is given by

$$\text{Var}(y) = \mu^2 \theta_1^2 + \delta_2^2 \sum_{j(2)=1}^{a_2} \mu_{j(2)}^2 + \delta_1^2 \sum_{k=1}^N \mu_k^2 \quad (14A.37)$$

where θ_1^2 is the variance component associated with the constant relative effect β'_1 . This model assumes that the accountability is based on the production measurements without correction of any kind, so that the control data are used only to provide estimates of θ_1^2 , δ_2^2 , and δ_1^2 .

When the first approach to correction is used, the variability in the corrected total is derived from Equation 14.29 as

$$\text{Var}(y) = \mu^2(\theta_3^2 + \delta_1^2/a_2 + \delta_2^2/a_2 + \delta_3^2/a_3) + \delta_2^2 \sum_{j(2)=1}^{a_2} \mu_{j(2)}^2 + \delta_1^2 \sum_{k=1}^N \mu_k^2$$

because the errors β'_i and $\bar{\beta}'_2$ have the same expectation β'_1 . The purpose of the average correction is to eliminate this constant effect, but the net result has been to replace the contribution from this effect to the total variability by a number of other (presumably smaller) variance components associated with the control measurements. The advantage in this procedure is that θ_1^2 , a component difficult to estimate, has been replaced by the presumably smaller quantity θ_3^2 and other components which are estimable from the control data. However, there is no assurance that the procedure will decrease the total variability, particularly when there is significant variation δ_2^2 from day to day in the constant daily effects.

Finally, if the second approach to correction is used and a daily factor is applied, the variability in the total is derived from Equation 14A.27 as

$$\text{Var}(y) = \mu^2 \theta_3^2 + \delta_1^2 \left(\sum_{j(2)=1}^{a_2} \mu_{j(2)}^2 + \sum_{k=1}^n \mu_k^2 \right) + \delta_3^2 \sum_{j(3)=1}^{a_3} \mu_{j(3)}^2$$

The net result of this daily correction has been to replace the day-to-day variability in the constant error effects by the variance in a single observation of the working standard and the variance components associated with the calibration of the working standards. This trade-off between the effect of systematic errors and the variance in determining corrections is a well-defined problem in connection with the treatment of measurement error.

REFERENCE

Ku, H. H. (Ed.), 1969, *Precision Measurement and Calibration*, NBS Special Publication 300, Vol. 1, U. S. Government Printing Office, Washington, D. C.

CHAPTER 15

Estimating and Controlling Measurement Variability

15.0 INTRODUCTION

In the preceding chapter models were developed for determining the effects of measurement errors on the results of the material accounting process. Procedures were established for determining the net effect of complex combinations of random error effects on the reported quantity of material in individual items, batches, strata, and the major components of the inventory difference. These procedures make it possible to express the expected value and variance of quantities such as inventory differences, shipper-receiver differences, and operator-inspector differences in terms of the corresponding characteristics of the error components. To proceed with the evaluation processes considered in Chapter 16, estimates of the measurement error parameters must be obtained. This chapter presents methods for obtaining such estimates. They represent a selection of those methods found to be useful in safeguards applications, rather than a comprehensive treatment of all possible techniques for estimating and controlling measurement error.

In practice, a wide variety of data sources are generally drawn upon in the course of estimating all of the measurement error parameters for a facility. The spectrum will range from data routinely taken in the facility for purposes not directly related to error estimation or the safeguarding of materials (e.g., process control data) to data from specially designed experiments or studies conducted expressly for the purposes of obtaining error estimates. Most of the data will normally come from three sources: (1) methods development and testing, including interlaboratory exchange programs; (2) the routine calibration of measurement systems, particularly those using nondestructive assay (NDA) methodology; and (3) continuing programs of measurement control for production measurements and laboratory analysis. Care must be taken to determine that the estimates obtained from these specialized sources remain valid under the conditions associated with routine use.

Neither the nature of the data sources nor the purposes of the analysis provide a completely appropriate basis for determining which statistical estimation and control procedures to present and the order in which to present them. Section 15.1 describes the calibration procedures required

for indirect measurement. Section 15.2 examines methods for the analysis of data from the routine testing and monitoring of direct measurements against known standards. The final section considers methods for estimating error variances from the replication of measurements on materials of unknown content under laboratory or production conditions.

15.1 CALIBRATION OF MEASUREMENT SYSTEMS

The techniques used in the calibration of measurement systems and in the control of routine measurements of concentration and bulk amounts are closely related. Calibration methods are largely based on the regression methodology of Chapter 6, while control techniques depend more on the hypothesis testing and estimation procedures described in Chapters 3, 4, and 5. Both applications depend to a large extent on the measurement of standards for which assigned weights, concentrations, or amounts of material have been established.

Calibration techniques relate a measured response, y , on an item to a standard value, x , of the result to be determined in some functional way; i.e., $y = f(x)$. The quantities y and x need not be in the same units; x might be in grams of ^{235}U and y in count rate for an NDA instrument. The process of calibration includes

1. Planning of the calibration data collection (number of standards, placement of standards within the range, number of replicated measurements on each standard, and number of repeated calibration runs).
2. Collection of the data.
3. Analysis of the calibration data.

Because the desired end result of the calibration process is to obtain estimates of the parameters appearing in $f(x)$, the data analysis problem is one of estimation and testing for parameter estimate stability over time or changing environmental conditions. In addition to the parameter estimates, various error variance estimates also should be obtained from the analysis of the calibration data. During the planning stage, sufficient replication of standards measurements and calibration runs should be incorporated to insure that all of the needed variances can be estimated.

Once the parameters have been estimated, the relationship is used to relate a response for a production item, observed on the y scale, to an estimated value on the x scale for that item. Statements about the standard error in this estimated value also are made, in part, on the basis of the calibration data.

15.1.1 Single-Point Calibration of NDA Instruments

A calibration and measurement problem frequently encountered when using NDA methods for materials control or safeguards activities can be

described as follows. Suppose the range of values for the characteristic of interest in the population to be measured is quite narrow. If a standard is available having an appropriate value for the particular calibration, use it for the population measurements. If such a standard is not available, it is common practice to randomly select from the population one (or perhaps a small number) of the process items and to make numerous, independent NDA measurements on that item. The NDA measurements on the selected "standard" item are interspersed among the measurements on the other process items. Once the NDA measurements on all the items have been completed, then the "standard" item is submitted for destructive analysis using a more reliable method, such as chemical or mass spectrometric analysis. In many cases, the uncertainty in the assigned value for the "standard" based on the destructive analysis results is quite small relative to the variability in repeated NDA measurements and almost can be ignored. This general type of calibration approach is sometimes referred to as the **single-point calibration** problem. Such single-point calibration data are analyzed assuming a linear calibration model with a known intercept at the origin. In the event that the characteristic in the population varies over too wide a range and that such a simple model may not be adequate, then more standards will be required to cover the calibration range and more complicated models may need to be considered for describing the calibration data.

The calibration model can be written as

$$y_i = \beta \mu + \epsilon_i$$

where

- y_i = the i^{th} measurement of the standard using NDA
- β = the slope of the calibration line
- μ = the true value of the characteristic for the standard
- ϵ_i = the error in the i^{th} NDA measurement; these errors are assumed to be independent with expected value zero and common variance σ^2

Let

- n = the number of NDA measurements of the standard
- \bar{y} = the mean of the measurements on the standard†
- μ_0 = the assigned value for the standard (may be based on destructive analyses); assumed to have expected value μ and variance σ_0^2

†As in Chapter 5, no distinction will be made in this chapter between the estimators \bar{Y} and S^2 of the sample mean and variance defined in Section 3.4.1 and the mean \bar{y} and variance s^2 of a particular set of observations. As is customary in applied texts, the lower case letters will be used.

Then, because $E(y_i) = E(\bar{y}) = \beta\mu$ and $E(\mu_0) = \mu$, the slope parameter, β , in this simple calibration model can be estimated by

$$\hat{\beta} = \frac{\bar{y}}{\mu_0} \quad (15.1)$$

The variance of this estimate is given approximately by (see Equation 2.127)

$$\text{Var}(\hat{\beta}) \cong \left(\frac{1}{\mu} \right)^2 \frac{\sigma^2}{n} + \left(\frac{\beta}{\mu} \right)^2 \sigma_0^2$$

A point estimate of this variance (see Equation 3.51) is given by

$$\text{V}\hat{\text{a}}\text{r}(\hat{\beta}) \cong \frac{s^2}{n\mu_0^2} + \frac{\bar{y}^2 \hat{\sigma}_0^2}{\mu_0^4} \quad (15.2)$$

where s^2 is the sample variance of the measurements on the standard and $\hat{\sigma}_0^2$ is an estimate of σ_0^2 . The latter estimate cannot be derived from the immediate calibration procedure, but must come from independent knowledge of the variability in the procedure by which the standard value was assigned.

If y_p is the measured value of y on a process item, an estimate of the value x_p of the characteristic is given by $\hat{x}_p = y_p/\hat{\beta}$. A simple extension of the process used to obtain Equation 15.2 leads to the approximation

$$\begin{aligned} \text{V}\hat{\text{a}}\text{r}(\hat{x}_p) &\cong \left(\frac{1}{\hat{\beta}} \right)^2 \sigma_p^2 + \left(\frac{y_p}{\hat{\beta}^2} \right)^2 \text{Var}(\hat{\beta}) \\ &\cong \left(\frac{1}{\hat{\beta}} \right)^2 \hat{\sigma}_p^2 + \left(\frac{y_p}{\hat{\beta}} \right)^2 \frac{s^2}{n\bar{y}^2} + \left(\frac{y_p}{\bar{y}} \right)^2 \hat{\sigma}_0^2 \end{aligned} \quad (15.3)$$

Again, an independent estimate of σ_p^2 is required unless it can be assumed that the measurement error variance on process items is the same as on the standard, in which case $\sigma_p^2 = \sigma^2$ and $\hat{\sigma}_p^2$ can be replaced by s^2 . For a single-point calibration, this assumption is tenable because the measurements generally are made using a fixed counting time and the standard is often a process item.

Technically, if the assigned value of the standard is a random variable whose range includes zero, then the variance of the estimate $\hat{\beta}$ given by Equation 15.1 will be infinite because there will be some nonzero probabil-

ity of a zero denominator. This situation is compounded in the case of the estimate \hat{x}_p because now the estimate $\hat{\beta}$ appears in the denominator of the estimator. Similar situations exist for other uses of calibration equations considered later in this chapter. In practice, this is seldom a concern for two reasons. The first reason is that a calibration and estimation procedure for which σ_0 is large compared to μ_0 , or for which $SE(\hat{\beta}) = \sqrt{\text{Var}(\hat{\beta})}$, the standard error† of $\hat{\beta}$, is large compared to β , is usually unacceptable. However, it is important to recognize that such situations are possible, especially with procedures involving working standards. The second reason is that the calibration and estimation process is not truly carried out as a single chance procedure, so that the range of values of μ_0 or $\hat{\beta}$ which would be accepted for subsequent use to estimate $\hat{\beta}$ or \hat{x}_p , respectively, are actually bounded away from zero.

Example 15.1 An active well coincidence counter operated in the fast mode is to be used to measure the ^{235}U content (in g) of 2.8% enriched UO_2 powder. The purpose of the measurement is to verify the assigned content of containers in one stratum of a plant inventory where all items are expected to have between 190 and 210 g of ^{235}U . There is no standard available, so a container is randomly selected from the inventory listing. The assigned value on the item tag for the selected container was 199.962 g of ^{235}U . Throughout the time period while the instrument was used to measure items from that stratum of the inventory, a total of 12 NDA measurements was made on the “standard” container. Listed below are the coincidence count rates for the measurements.

**Real Coincidence Count Rate Measurement
Results on Standard (y)**

| | |
|--------|--------|
| 34.112 | 33.773 |
| 32.891 | 32.797 |
| 33.581 | 32.225 |
| 32.946 | 33.631 |
| 31.312 | 33.650 |
| 33.355 | 33.280 |

†In the rest of this chapter and in subsequent chapters, the term standard error, or standard error of estimate, is used to denote the square root of the variance, or standard deviation, of the point estimator of a parameter (see Section 3.4). This emphasizes the fact that the variances derived in Chapter 14 for accounting estimates, and in this chapter for the parameters of measurement systems, are frequently complex functions of the variances of one or more random variables. The estimates may or may not involve replicate observations on these random variables, and frequently the component variances are not estimable from the data on which the estimate is based.

For these measurement results

$$n = 12$$

$$\bar{y} = 33.1294$$

$$s^2 = 0.5911$$

$$s = 0.7688$$

After all the NDA measurements were completed, the gross and net weights of the powder in the selected container were determined, and five random samples of the powder were selected and sent to the laboratory for destructive analysis. The assigned value for the standard was calculated using

$$\mu_0 = W \bar{C} \bar{I}$$

where

W = net weight of the powder in the selected container

\bar{C} = the average concentration of uranium determined from the five destructive analyses

\bar{I} = the average weight fraction of ^{235}U in the uranium

The actual assigned value obtained as a result of the weighing and analytical measurements was $\mu_0 = 200.799 \text{ g } ^{235}\text{U}$ and the standard deviation associated with the assigned value was estimated as $\hat{\sigma}_0 = 0.5395 \text{ g}$.

From these results, the estimate $\hat{\beta}$ is calculated from Equation 15.1 as

$$\hat{\beta} = \frac{33.1294}{200.799} = 0.164988$$

From Equation 15.2, an estimate of the variance of $\hat{\beta}$ is

$$\begin{aligned} \text{Var}(\hat{\beta}) &\cong \left(\frac{1}{200.799} \right)^2 \frac{(0.5911)}{12} + \left(\frac{33.1294}{200.799^2} \right)^2 (0.5395)^2 \\ &= 1.418178 \times 10^{-6} \end{aligned}$$

and the estimated standard error of $\hat{\beta}$ is

$$\hat{\text{SE}}(\hat{\beta}) = \sqrt{\text{Var}(\hat{\beta})} = 0.001191$$

For a container from inventory counted a single time (using the same counting time used in the calibration) with measured coincidence count rate of 31.376, the calculated ^{235}U content of the container is

$$\hat{x}_p = \frac{31.376}{0.164988} = 190.171 \text{ g}$$

and the variance and standard error of that calculated value are estimated as

$$\begin{aligned} \text{Vâr}(\hat{x}_p) &= \left(\frac{1}{0.164988} \right)^2 (0.5911) \\ &\quad + \left(\frac{31.376}{0.164988} \right)^2 \frac{(0.5911)}{12(33.1294)^2} \\ &\quad + \left(\frac{31.376}{33.1294} \right)^2 (0.5395)^2 \\ &= 21.7148 + 1.6231 + 0.2611 \\ &= 23.5990 \text{ (g)}^2 \end{aligned}$$

and

$$\hat{\text{SÊ}}(\hat{x}_p) = 4.858 \text{ g}$$

15.1.2 General Linear Calibration

A first extension of the single-point calibration in the previous section is the case where the calibration relationship is linear and it cannot be assumed that the line goes through the origin. This unrestricted linear calibration model is used for many measurement systems. The following sections consider the procedures for general linear calibration under several different assumptions.

15.1.2.1 Constant Variance in the Measured Responses

This section considers general linear calibration when the y variable has constant variance over the range of x of interest. The calibration data are assumed to consist of pairs of (x_i, y_i) observations, where the x_i are fixed values of the i^{th} standard, assumed to be known with negligible error.

At any fixed value of x , y is distributed with mean value on the calibration line and with unknown, but constant, variance. In many practical applications it is reasonable to also assume that y is normally distributed.

The model for general linear calibration with constant variance is

$$y_{ij} = \beta_0 + \beta_1 \mu_i + \epsilon_{ij} \quad (15.4)$$

where

- $i = 1, 2, \dots, m$ standards
- $j = 1, 2, \dots, n$ replicate measurements on the i^{th} standard
- y_{ij} = j^{th} measurement of the i^{th} standard
- μ_i = assigned value of the i^{th} standard; usually assumed to be known exactly but may have assigned variances
- β_0 = intercept parameter
- β_1 = slope parameter
- ϵ_{ij} = error associated with the j^{th} measurement of the i^{th} standard; these errors are assumed to be independent with expected value zero and common variance σ^2

Methods for calculating all of the required parameter estimates (intercept, slope, and common variance) can be found in Sections 6.2.3 and 6.4.1.

There are two cases frequently encountered when a linear calibration is used to estimate the nuclear material content of items or containers. In the first case, a measurement is required at only a single point along the calibration line to provide the desired material content estimate. An example might involve an NDA instrument used to measure the ^{235}U content of reactor fuel pellets. As long as background radiation during a production measurement is controlled at the same (ideally, very low) level as was present during calibration, then only a measurement with the unknown pellet in the instrument is required and a background measurement is not needed. In the second case, measurements at two separated points on the calibration line are required to obtain the desired material content estimate. An example is the determination of the volume of liquid transferred from an accountability tank. To determine the net amount transferred, measurements with the tank full of liquid and after the transfer must both be transformed to amounts of liquid using the calibration equation.

For the first case, if y_p is the measurement system response on a process item, then the value of the characteristic for the item can be estimated using

$$\hat{x}_p = (y_p - \hat{\beta}_0) / \hat{\beta}_1 \quad (15.5)$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates of the two calibration parameters. Once

estimated from a given set of calibration data, these estimates are fixed and will produce the same error effect due to calibration in all item results having the same y_p values. Using Equation 3.51, the variance of \hat{x}_p can be approximated by

$$\begin{aligned} \text{Vâr}(\hat{x}_p) &\cong \frac{\text{Vâr}(y_p)}{\hat{\beta}_1^2} + \left(\frac{1}{\hat{\beta}_1} \right)^2 \text{Vâr}(\hat{\beta}_0) \\ &\quad + \left(\frac{y_p - \hat{\beta}_0}{\hat{\beta}_1^2} \right)^2 \text{Vâr}(\hat{\beta}_1) \\ &\quad + \frac{2(y_p - \hat{\beta}_0)}{\hat{\beta}_1^3} \text{Côv}(\hat{\beta}_0, \hat{\beta}_1) \end{aligned} \quad (15.6)$$

where

$\text{Vâr}(y_p)$ = estimated variance of the measured response for the process item

$\text{Vâr}(\hat{\beta}_0)$ = estimated variance of the intercept estimate

$\text{Vâr}(\hat{\beta}_1)$ = estimated variance of the slope estimate

$\text{Côv}(\hat{\beta}_0, \hat{\beta}_1)$ = estimated covariance between intercept and slope estimates

The symbol $\text{Vâr}(y_p)$ is used for the estimated variance of the measurement response for the process item to emphasize the fact that the random measurement variance may be different during calibration than in routine measurement of process items. From a regression analysis of the calibration data, the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\text{Vâr}(\hat{\beta}_0)$, $\text{Vâr}(\hat{\beta}_1)$, $\text{Côv}(\hat{\beta}_0, \hat{\beta}_1)$, and the residual error variance are generally available (Sections 6.2.3 and 6.4.3). The residual mean square from the regression analysis (Section 6.4.1) may or may not be a good estimate of $\text{Var}(y_p)$. During calibration, more care may have been taken than is usual when making production measurements, or more experienced personnel may have been used. Generally, an estimate for $\text{Var}(y_p)$ should be obtained from current data from independent replicate measurements on part (or all) of the process items measured. If the residual variance estimate from the regression analysis is going to be used for $\text{Vâr}(y_p)$, it is desirable to compare an independent estimate with the residual variance (Section 4.3.3) to verify that it is, in fact, a valid estimate for the measurement of process items.

For the second case, the notation y_u will be used for the upper measurement response (height of liquid in tank before transfer) and y_l for the

lower measurement response (height after the transfer). For the linear model being discussed in this section, an estimate of the net material transferred is given by

$$\Delta\hat{x} = \frac{(y_u - \hat{\beta}_0) - (y_l - \hat{\beta}_0)}{\hat{\beta}_1} = \frac{y_u - y_l}{\hat{\beta}_1} \quad (15.7)$$

Assuming that the variances for both the upper and lower measurement responses are equal, and using $\text{V}\hat{\text{a}}\text{r}(y)$ as the symbol for the estimate of that variance, the variance of $\Delta\hat{x}$ can be approximated by

$$\text{Var}(\Delta\hat{x}) \cong 2 \left(\frac{1}{\hat{\beta}_1} \right)^2 \text{V}\hat{\text{a}}\text{r}(y) + \left(\frac{y_u - y_l}{\hat{\beta}_1^2} \right)^2 \text{V}\hat{\text{a}}\text{r}(\hat{\beta}_1) \quad (15.8)$$

Example 15.2 The scale calibration data in Table 15.1 were generated from information given in an IAEA manual (IAEA, 1980). One use of the scale is the determination of the net weight of nuclear material received at a facility by measuring the gross weight of a full container, measuring the tare weight of the container with the same scale, and obtaining the net weight estimate as the difference in these two measured values. For this example, it is assumed that the calibration over the range of intended use is linear and that the standard deviation of measurement results throughout the range is constant.

Two standards were selected for the calibration; one with an assigned value of 8878.0 g (corresponding to the gross weight of received containers) and one with an assigned value of 1591.7 g (corresponding to the weight of empty containers). For this example, the assigned values are assumed to be known without error. Each standard was weighed 10 times. The order of weighing was randomized and the standard weighing procedure followed. The 20 (x_i, y_i) data pairs for the calibration measurements are presented in the following table. The analysis will treat these data as 20 independent pairs. Because there are only two standard values, they could have been analyzed as a two-point calibration with $n_1 = n_2 = 10$ replications at each of the $m = 2$ assigned values of the standards (Section 6.4.4). The residual variance in this case is simply the pooled variance about the mean for each standard and estimates only measurement variability. No test of the assumption of a linear fit to the two points (x_1, \bar{y}_1) and (x_2, \bar{y}_2) is possible.

TABLE 15.1
Scale Calibration Data

| Observation | Assigned value
of standard x_i (g) | Measured scale
response y_i (g) |
|-------------|---|--------------------------------------|
| 1 | 8878.0 | 8882.1 |
| 2 | 1591.7 | 1589.2 |
| 3 | 1591.7 | 1587.1 |
| 4 | 1591.7 | 1593.1 |
| 5 | 8878.0 | 8880.4 |
| 6 | 1591.7 | 1588.4 (1688.4)† |
| 7 | 1591.7 | 1584.5 |
| 8 | 8878.0 | 8876.3 |
| 9 | 8878.0 | 8893.6 |
| 10 | 1591.7 | 1594.0 |
| 11 | 8878.0 | 8887.0 |
| 12 | 1591.7 | 1592.4 |
| 13 | 1591.7 | 1587.5 |
| 14 | 1591.7 | 1583.8 |
| 15 | 8878.0 | 8884.1 |
| 16 | 8878.0 | 8874.9 |
| 17 | 1591.7 | 1582.5 |
| 18 | 8878.0 | 8883.8 |
| 19 | 8878.0 | 8876.7 |
| 20 | 8878.0 | 8888.6 |

†This second value will be used in a later part of the example.

For a linear regression analysis, the following calculated quantities are needed:

$$n = 20$$

$$\sum x_i = 104,697$$

$$\sum x_i^2 = 813,523,928.9$$

$$\sum y_i = 104,710$$

$$\sum x_i y_i = 813,890,720.25$$

$$\sum y_i^2 = 814,258,318.5$$

The estimated values for the slope and intercept are

$$\begin{aligned}
 \hat{\beta}_1 &= \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \\
 &= \frac{20(813,890,720.25) - (104,697)(10,470)}{20(813,523,928.9) - (104,697)^2} \\
 &= 1.00112539972 = 1.0011254 \\
 \hat{\beta}_0 &= \frac{\sum y}{n} - \hat{\beta}_1 \frac{\sum x}{n} = \frac{104,710}{20} - 1.00112539972 \left(\frac{104,697}{20} \right) \\
 &= -5.24129872
 \end{aligned}$$

The residual error variance estimate is

$$\begin{aligned}
 s^2 &= \frac{1}{n-2} \left\{ \sum y_i^2 - \frac{(\sum y_i)^2}{n} - \hat{\beta}_1^2 \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \right\} \\
 &= \frac{1}{18} \left\{ 814,258,318.5 - \frac{(104,710)^2}{20} \right. \\
 &\quad \left. - (1.00112539972)^2 \left[813,523,928.9 - \frac{(104,697)^2}{20} \right] \right\} \\
 &= \frac{1}{18} (462.252) = 25.6807 \text{ (g)}^2
 \end{aligned}$$

The estimated variances and covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$\begin{aligned}
 \text{Var}(\hat{\beta}_0) &= s^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right] \\
 &= \frac{s^2}{n} \left[1 + \frac{(\sum x_i)^2}{n \sum x_i^2 - (\sum x_i)^2} \right]
 \end{aligned}$$

$$= \frac{25.6807}{20} \left[1 + \frac{(104,697)^2}{20(813,523,928.9) - (104,697)^2} \right]$$

$$= 3.9352$$

$$\text{Var}(\hat{\beta}_1) = s^2 \left[\frac{1}{\sum(x_i - \bar{x})^2} \right] = \left[\frac{20(25.68066667)}{20(813,523,928.9) - (104,697)^2} \right]$$

$$= 9.6744 \times 10^{-8}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = s^2 \left[\frac{-\bar{x}}{\sum(x_i - \bar{x})^2} \right] = s^2 \left[\frac{-\sum x_i}{n \sum x_i^2 - \sum(x_i)^2} \right]$$

$$= \frac{-(25.68066667)(104,697)}{20(813,523,928.9) - (104,697)^2}$$

$$= -5.0644 \times 10^{-4}$$

To demonstrate the use of this scale calibration equation, suppose that a container of material is received for which the measured gross weight is 8875.2 g and the measured weight of the empty container is 1583.9 g. Based on these measurements, the estimated amount of material received is calculated from Equation 15.7 as

$$\Delta \hat{x} = \frac{8875.2 - 1583.9}{1.0011254} = 7283.1$$

If it is reasonable to use 25.6807 as the variance estimate for measuring both the gross and tare weights of process items, then from Equation 15.8

$$\begin{aligned} \text{Var}(\Delta \hat{x}) &= 2 \left(\frac{1}{1.0011254} \right)^2 (25.6807) \\ &\quad + \left(\frac{7291.3}{(1.0011254)^2} \right)^2 (9.6744 \times 10^{-8}) \\ &= 51.2466 + 5.1201 = 56.3667 (\text{g})^2 \end{aligned}$$

and the estimated standard error in the estimated amount of material received is 7.51 g.

The residuals r_i calculated using Equation 6.31 and the standardized residuals r_i^* calculated using Equation 6.32 are presented in Table 15.2 (see Section 10.5). The actual measured responses appear in the table of calibration data. The predicted responses for the two standards are 1588.25 and 8882.75 g; in this case, they are the mean values of the 10 measurements on each standard.

TABLE 15.2
Scale Calibration Residuals
and Standardized Residuals

| Observation | r_i | r_i^* |
|-------------|-------|---------|
| 1 | -0.65 | -0.135 |
| 2 | 0.95 | 0.198 |
| 3 | -1.15 | -0.239 |
| 4 | 4.85 | 1.009 |
| 5 | -2.35 | -0.489 |
| 6 | 0.15 | 0.031 |
| 7 | -3.75 | -0.780 |
| 8 | -6.45 | -1.342 |
| 9 | 10.85 | 2.257 |
| 10 | 5.75 | 1.196 |
| 11 | 4.25 | 0.884 |
| 12 | 4.15 | 0.863 |
| 13 | -0.75 | -0.156 |
| 14 | -4.45 | -0.926 |
| 15 | 1.35 | 0.281 |
| 16 | -7.85 | -1.633 |
| 17 | -5.75 | -1.196 |
| 18 | 1.05 | 0.218 |
| 19 | -6.05 | -1.258 |
| 20 | 5.85 | 1.217 |

The h_{ii} value (see Section 6.4.5.2) for the observations on the low standard with the common x_i value 1591.7 was calculated as

$$(1 \ 1591.7) \begin{pmatrix} 0.153234 & -1.972060 \times 10^{-5} \\ -1.972060 \times 10^{-5} & 3.767176 \times 10^{-9} \end{pmatrix} \begin{pmatrix} 1 \\ 1591.7 \end{pmatrix} = 0.1000$$

Note that the numerical values in the $(X'X)^{-1}$ matrix are exactly the quantities in brackets used to calculate the estimates $\hat{V}\hat{a}r(\hat{\beta}_0)$, $\hat{V}\hat{a}r(\hat{\beta}_1)$, and $\hat{C}\hat{o}v(\hat{\beta}_0, \hat{\beta}_1)$ from s^2 (see Section 6.4.3). The h_{ii} values for the observations on the high standard are also 0.1, which reflects the two-point nature of

the calibration. Plots of the standardized residuals in order of measurement and for the two assigned standard values are shown in Figure 15.1. Residuals, standardized residuals, and residual plots are usually provided as part of the generalized regression routines contained in the statistical software packages available for computers.

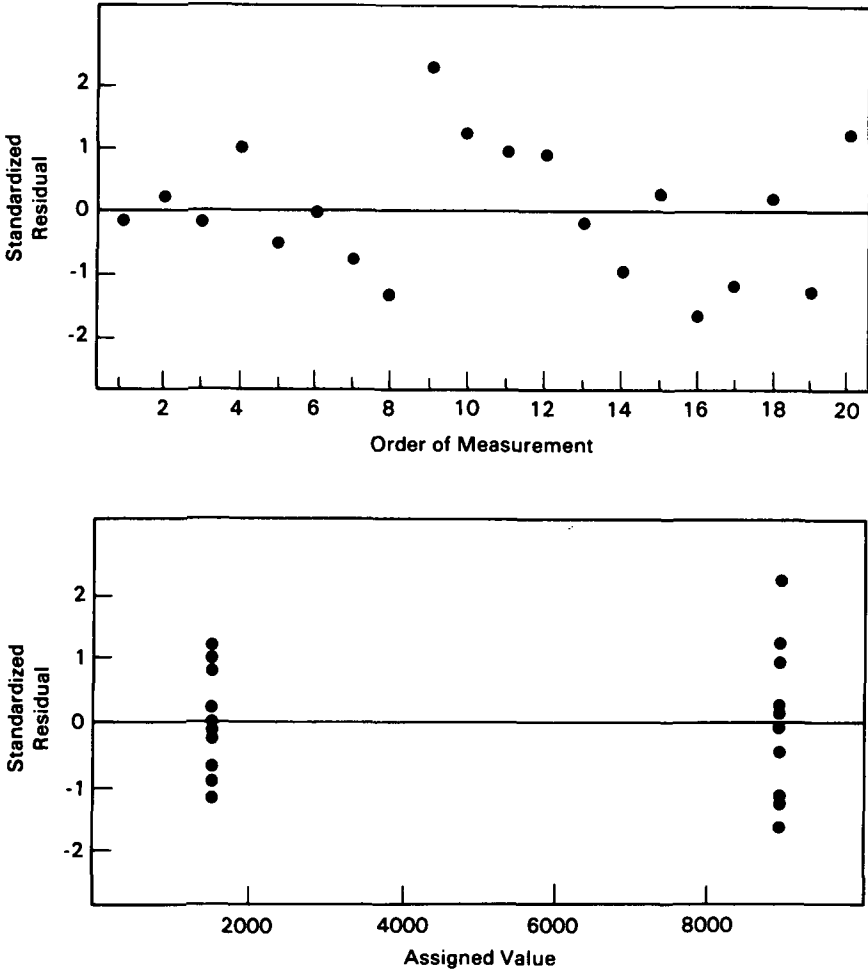


Figure 15.1 Residual plots for scale calibration.

The r_i^* value for Observation 9 is somewhat high when compared with the t value of 2.101 for 18 degrees of freedom and $\alpha = 0.05$. However, by chance alone it would be expected that about 5% of the r_i^* values would be larger than this value. There does not appear to be sufficient evidence to justify the deletion of Observation 9 as an abnormal result. To

examine the effect of an abnormal result on the analysis and the nature of the r_i^* values with one outlier present, consider the same set of data with the y_i value appearing in parentheses as the sixth observation. This could have occurred if the 5 in 1588.4 had been misrecorded as a 6. All of the detailed calculations for this second part of the example will not be shown. The parameter estimates for the second analysis are

$$\hat{\beta}_0 = 6.943212$$

$$\hat{\beta}_1 = 0.999753$$

$$s^2 = 527.347233$$

$$\text{Vâr}(\hat{\beta}_0) = 80.807730$$

$$\text{Vâr}(\hat{\beta}_1) = 1.986610 \times 10^{-6}$$

$$\text{Côv}(\hat{\beta}_0, \hat{\beta}_1) = 1.039960 \times 10^{-2}$$

The standardized residuals were calculated using the results from the analysis of the data set with one known outlier present and are presented in Table 15.3. Because the $(X'X)^{-1}$ matrix remains the same for analysis

TABLE 15.3
Standardized Residuals
Data Set with One Outlier

| Observation | r_i | r_i^* |
|-------------|--------|---------|
| 1 | -0.65 | -0.030 |
| 2 | -9.05 | -0.415 |
| 3 | -11.15 | -0.512 |
| 4 | -5.15 | -0.236 |
| 5 | -2.35 | -0.108 |
| 6 | 90.15 | 4.138 |
| 7 | -13.75 | -0.631 |
| 8 | -6.45 | -0.296 |
| 9 | 10.85 | 0.498 |
| 10 | -4.25 | -0.195 |
| 11 | 4.25 | 0.195 |
| 12 | -5.85 | -0.269 |
| 13 | -10.75 | -0.493 |
| 14 | -14.45 | -0.663 |
| 15 | 1.35 | 0.062 |
| 16 | -7.85 | -0.360 |
| 17 | -15.75 | -0.723 |
| 18 | 1.05 | 0.048 |
| 19 | -6.05 | -0.278 |
| 20 | 5.85 | 0.269 |

of the revised data set, the h_{ii} values are the same for the calculation of these new standardized residuals.

Because the recording error was large, 1688.4 when it should have been 1588.4, the r_i^* value for the observation stands out clearly. With 11 of the other r_i^* values having absolute values less than 1.0, the value greater than 4.1 is obviously different. This can be seen clearly on the residual plots of Figure 15.2 compared to Figure 15.1.

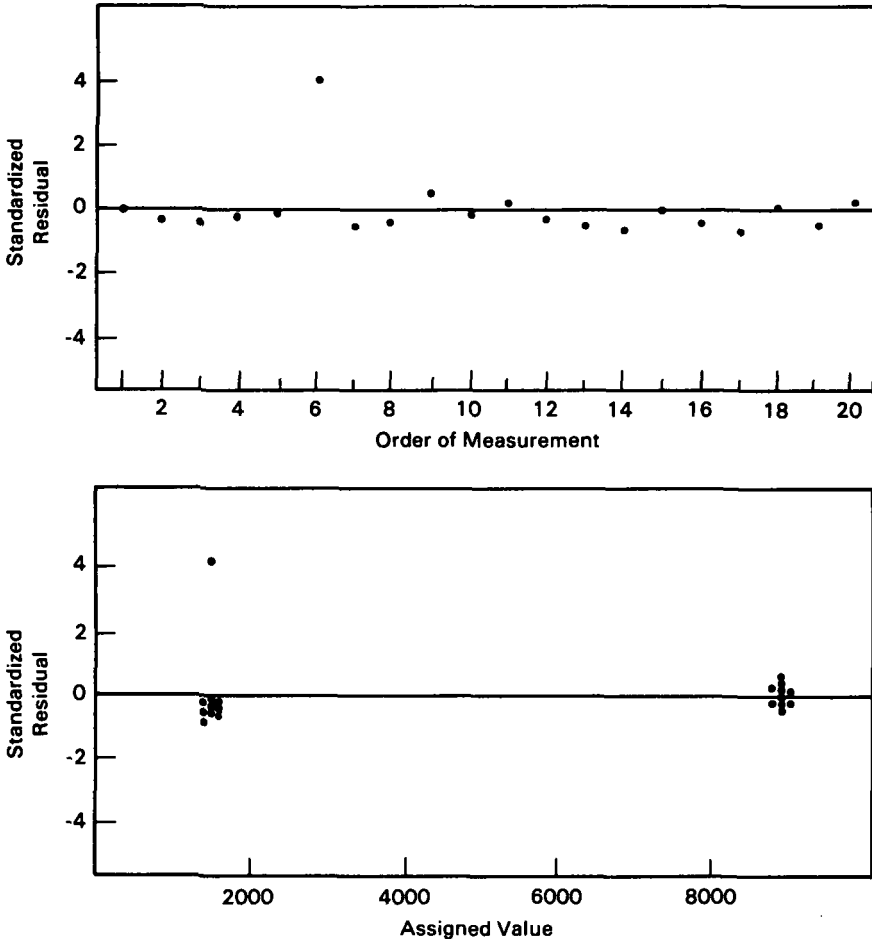


Figure 15.2 Residual plots for scale calibration with one outlier present.

15.1.2.2 Nonhomogeneity of Variance in the Responses

In linear calibration applications for which the variance of the y variable is not constant over the range of x of interest, the analysis of the data

to obtain parameter estimates (intercept, slope, and variances) can be accomplished as long as the variances are known for the x values used. In fact, the data can be analyzed when only the relative magnitudes of the variances are known, e.g., the variance at \bar{x}_2 is 1.5 times the variance at \bar{x}_1 , the variance at \bar{x}_3 is 1.75 times the variance at \bar{x}_1 , etc. The problem of calibration when the variance of the instrument (or measurement system) response is not constant is encountered quite frequently for NDA measurement methods where the response is related to the amount of the element (or isotope) of interest. It is sometimes possible to adjust such variables as counting times during the calibration data gathering runs so the resulting data have equal variances over the range. Such adjustments require pre-knowledge about the measurement system behavior, however, and this knowledge must be used at the data collection planning stage.

The following example illustrates the weighted regression analysis which should be used to obtain the calibration parameter estimates when the measurement response variance is not constant throughout the calibration range.

Example 15.3 Assume that a high-level neutron coincidence counter (HLNCC) was calibrated for neutron-multiplication-corrected measurements of the mass of ^{240}Pu effective contained in PuO_2 samples. Three standards were used for the calibration. Replicate 1000 s measurements were made on the low and high standards and a single measurement on the intermediate standard. The expected variance in the corrected real rate was determined from expected counting rates for each standard. The data have been synthesized from some HLNCC results in Krick (1980). The results are given in Table 15.4.

Over the limited range of ^{240}Pu effective mass used for this calibration exercise, it is expected that the corrected real rate \hat{R} will be linearly related to the ^{240}Pu effective mass. Because a constant counting time of 1000 s was used for all measurements, however, the variance of the response would not be expected to be constant. The variances of the responses due to the coincidence counting procedure are given in the fourth column of Table 15.4. The weights $w_i = 1/\sigma_i^2$ are given in the last column.

Using the methods presented in Section 6.3.2, Case 2, the weighted least squares estimates were calculated. The model used is the same as the one given in Section 15.1.2.1, with the important exception that $\text{Var}(\epsilon_{ij}) = \sigma_i^2$ is assumed to change from response to response (Equation 6.17). The calculations for this case are identical to those for unweighted linear regression, except that weighted rather than unweighted sums and sums of

TABLE 15.4
Calibration Data for a HLNCC Instrument
Using PuO₂ Standards

| Observation | ²⁴⁰ Pu eff
(g) | \dot{R}
(counts/s) | σ_i^2 | w_i |
|-------------|------------------------------|-------------------------|--------------|-----------|
| 1 | 97.45 | 531.26 | 63.372 | 0.0157798 |
| 2 | 97.45 | 536.71 | 63.372 | 0.0157798 |
| 3 | 584.3 | 3157.01 | 2221.81 | 0.0004501 |
| 4 | 97.45 | 551.16 | 63.372 | 0.0157798 |
| 5 | 292.3 | 1613.08 | 573.60 | 0.0017434 |
| 6 | 584.3 | 3109.82 | 2221.81 | 0.0004501 |
| 7 | 584.3 | 3204.48 | 2221.81 | 0.0004501 |
| 8 | 584.3 | 3133.61 | 2221.81 | 0.0004501 |
| 9 | 584.3 | 3205.44 | 2221.81 | 0.0004501 |
| 10 | 584.3 | 3173.29 | 2221.81 | 0.0004501 |
| 11 | 584.3 | 3072.59 | 2221.81 | 0.0004501 |
| 12 | 97.45 | 531.59 | 63.372 | 0.0157798 |

squares of standard values and responses are used. The corresponding sums are shown in the following table.

| Unweighted regression | Weighted regression |
|-----------------------|---------------------|
| n | $\sum w_i$ |
| $\sum x_i$ | $\sum w_i x_i$ |
| $\sum x_i^2$ | $\sum w_i x_i^2$ |
| $\sum y_i$ | $\sum w_i y_i$ |
| $\sum x_i y_i$ | $\sum w_i x_i y_i$ |
| $\sum y_i^2$ | $\sum w_i y_i^2$ |

For the data of this example, with $x =$ ²⁴⁰Pu effective and $y = \dot{R}$

$$\sum w_i = 0.0680133$$

$$\sum w_i x_i = 8.50151587$$

$$\sum w_i x_i^2 = 1824.035927$$

$$\sum w_i y_i = 46.67768875$$

$$\sum w_i x_i y_i = 9929.916456$$

$$\sum w_i y_i^2 = 54,075.28424$$

The following estimates are obtained:

$$\hat{\beta}_0 = 13.9513$$

$$\hat{\beta}_1 = 5.3789$$

$$s^2 = 1.2029$$

$$\text{V}\hat{\text{a}}\text{r}(\hat{\beta}_0) = 42.3719$$

$$\text{V}\hat{\text{a}}\text{r}(\hat{\beta}_1) = 0.00157993$$

$$\text{C}\hat{\text{o}}\text{v}(\hat{\beta}_0, \hat{\beta}_1) = -0.197488$$

The linear calibration equation can be used to measure process containers by solving for x . If y_p is the neutron-multiplication-corrected response on a process item, then the predicted ^{240}Pu effective mass, \hat{x}_p is

$$\hat{x}_p = \frac{y_p - \hat{\beta}_0}{\hat{\beta}_1} = \frac{y_p - 13.9513}{5.3789}$$

Suppose that two containers of plutonium oxide powder were received in a material balance area and that 1000 s measurements provided y_p values of 2172.87 and 2057.22. The same background information that provided the measurement variances for weighting of the calibration data indicated that these two measurements should have variances of 1250 and 1100, respectively. The sum of the grams ^{240}Pu effective and the approximate variance of this sum are needed. The sum can be estimated as

$$\hat{x}_{\text{sum}} = \frac{(y_{p1} + y_{p2}) - 2\hat{\beta}_0}{\hat{\beta}_1} = \frac{4230.09 - 2(13.9513)}{5.3789} = 781.24 \text{ g}$$

The variance of the sum is broken down into two components, the variance associated with the new measurements and the variance due to the uncertainty in the estimated calibration parameters. The component arising from the new measurements can be estimated as

$$\text{V}\hat{\text{a}}\text{r}_r(\hat{x}_{\text{sum}}) \cong \frac{1250 + 1100}{(5.3789)^2} = 81.2234 \text{ (g)}^2$$

and the component due to calibration uncertainty as

$$\text{V}\hat{\text{a}}\text{r}_c(\hat{x}_{\text{sum}}) \cong \left[-\frac{y_{p1} + y_{p2} - 2\hat{\beta}_0}{\hat{\beta}_1^2} \right]^2 \text{V}\hat{\text{a}}\text{r}(\hat{\beta}_1) + \left[\frac{-2}{\hat{\beta}_1} \right]^2 \text{V}\hat{\text{a}}\text{r}(\hat{\beta}_0)$$

$$\begin{aligned}
& + 2 \left[\frac{-2}{\hat{\beta}_1} \right] \left[-\frac{y_{p1} + y_{p2} - 2\hat{\beta}_0}{\hat{\beta}_1^2} \right] \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\
& = \frac{(4202.19)^2}{(5.3789)^4} (0.00157993) + \left[\frac{2}{5.3789} \right]^2 (42.3719) \\
& \quad + \frac{4(4202.19)}{(5.3789)^3} (-0.197488) \\
& = 33.3285 + 5.8580 - 21.3302 \\
& = 17.8563 \text{ (g)}^2
\end{aligned}$$

A measure of the total uncertainty is the sum of these two variance components, 99.0797 (g)², and the estimated standard error is 9.95 g or 1.27% of the estimated sum.

Both the residuals calculated using Equation 6.31 and the standardized residuals calculated using an equation analogous to Equation 6.32 for the weighted regression case are presented in Table 15.5.

TABLE 15.5
Data and Residuals for HLNCC Calibration

| Observation | x_i | y_i | \hat{y}_i | r_i | r_i^* |
|-------------|-------|---------|-------------|---------|---------|
| 1 | 97.45 | 531.26 | 538.125 | -6.865 | -0.994 |
| 2 | 97.45 | 536.71 | 538.125 | -1.415 | -0.204 |
| 3 | 584.3 | 3157.01 | 3156.843 | 0.167 | 0.004 |
| 4 | 97.45 | 551.16 | 538.125 | 13.035 | 1.888 |
| 5 | 292.3 | 1613.08 | 1586.204 | 26.876 | 1.176 |
| 6 | 584.3 | 3109.82 | 3156.843 | -47.023 | -1.070 |
| 7 | 584.3 | 3204.48 | 3156.843 | 47.637 | 1.084 |
| 8 | 584.3 | 3133.61 | 3156.843 | -23.233 | -0.529 |
| 9 | 584.3 | 3205.44 | 3156.843 | 48.597 | 1.106 |
| 10 | 584.3 | 3173.29 | 3156.843 | 16.447 | 0.374 |
| 11 | 584.3 | 3072.59 | 3156.843 | -84.253 | -1.918 |
| 12 | 97.45 | 531.59 | 538.125 | -6.535 | -0.946 |

Because the residual error standard deviation from the weighted regression analysis is very close to the unit variance that would be expected if the individual error variance used to obtain the weights were correctly estimated, the standardized residuals were computed by a modification of Equation 6.32 given by

$$r_i^* = \frac{r_i}{\sigma_i^2 \sqrt{1 - h_{ii}}}$$

$$= \frac{r_i}{\sqrt{\sigma_i^2 - \text{Var}(\hat{y}_i)}}$$

where σ_i^2 is the variance of the standard measurement for the i^{th} observation, and $\text{Var}(\hat{y}_i)$ is the variance of the predicted value, \hat{y}_i , for the i^{th} observation. While it is generally of interest to examine all residuals and standardized residuals to detect possible outlier observations and to check for lack of fit of the chosen model, in this example only the single measurement on the intermediate standard ($x_i = 292.3$) provides information concerning lack of fit. This limited amount of information from Observation 5 does not indicate lack of fit, because the r^* value is only slightly larger than 1.0 (for $\alpha = 0.10$, $t_{0.95}(10) = 1.812$). Because none of the r^* values are large, there is no indication of outliers in the data set.

15.1.2.3 Recalibration Procedures

All measurement systems need to be recalibrated on a routine schedule because of the difficulties in assuring a stable measurement system over extended periods of time. The need for frequent recalibration is especially true for some of the NDA measurement methods. If the calibration relationship changes or shifts, it is important to understand the nature of the changes or shifts, and to treat the calibration data in an appropriate way to select the best estimates for the calibration parameters and the uncertainty in the measurement of process items.

In this section, it is assumed that a linear model with unknown intercept is adequate. The methods discussed can be applied to either the constant variance or the nonconstant variance cases as long as sufficient information is available in the nonconstant variance case to allow weighted regression analyses to be performed. In most practical applications, the calibration equation actually applied is the one based on the most recent set of data. There may be quite a few ways in which the calibration parameters could change over time. It could be that the parameters remain constant within a time period and that random changes occur at the end of each period. On the other hand, it could be that the parameters change continuously over time. For this second type of change, the actual changes might follow some smooth time series or the changes might occur erratically or randomly as time progresses. An extreme example of such a situation might be where a recalibration within a matter of hours following a previous calibration would exhibit changes in the parameter as large as

those which would occur if a longer time had been allowed to elapse between calibration runs.

Example 15.4 In calibrating an NDA system for the analysis of uranium solid waste barrels based on gamma counting, the model $y = \beta_0 + \beta_1 x$ is assumed, where y is the net counts per 100 s and x is the grams of ^{235}U in standard barrels. It is assumed that sufficient experience has been had with the system so that the standard deviation, σ_i , for the measurement of the i^{th} standard is known precisely for 100 s counts. Data from the five most recent calibration runs are presented in Table 15.6. For each run, six standard waste barrels were used and two measurements were performed on each standard. The order of measurement was randomized. Run 5 is the current calibration run. From these data the results in Table 15.7 were obtained using weighted regression analysis of each of these five sets and the combined data.

To check for changes in the calibration equation among the runs, the analysis of variance shown in Table 15.8 was performed. In addition to testing for differences among the runs, this analysis includes a test of the adequacy of the linear calibration function. All of the quantities necessary to construct the table can be obtained from analyses of variance for the individual and combined regressions (Section 6.4.1) except the weighted sum of squares of differences between replicates needed to obtain the sum of squares "between replicate measurements." Specifically, the entries in the table were obtained as follows:

1. The sum of squares due to the **common regression** is the sum of squares explained by the weighted fit of a single calibration line to the combined data from the five runs (see Equation 6.18 and Table 6.2). The residual sum of squares from this combined fit is the combination of the "residual from regression" and the sum of squares explained by the differences "between intercepts" and "between slopes." Note that $(45.7356 + 22.9745 + 49.1637)/(4 + 4 + 50) = 2.0323$, the residual error variance estimate previously obtained for the combined fit.

2. The **residual from regression** is the combined residual sum of squares from the weighted fit of a straight line to each individual run, which has a combined total of $5(12 - 2) = 50$ degrees of freedom. The **total regression** is obtained by the difference between this combined residual and the total sum of squares.

3. The **between intercepts** sum of squares is obtained as the difference between the total sum of squares for the combined runs and the sum of the total sums of squares within each run.

4. The **between slopes** sum of squares is the difference between that explained by "total regression" and the amounts explained by the "com-

TABLE 15.6
Calibration of Uranium Solid Waste Barrel System

| x_i | σ_i | w_i | Run 1 | | Run 2 | | Run 3 | | Run 4 | | Run 5 | |
|-------|------------|------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | | y_i | y_i | y_i | y_i | y_i | y_i | y_i | y_i | y_i | y_i |
| 1 | 410 | 5.949×10^{-6} | 3,210 | 3,639 | 4,240 | 4,661 | 4,137 | 4,095 | 3,481 | 3,269 | 4,123 | 3,581 |
| 4 | 485 | 4.251×10^{-6} | 11,062 | 10,290 | 10,971 | 10,507 | 11,895 | 12,307 | 11,706 | 10,830 | 11,255 | 11,352 |
| 7 | 550 | 3.306×10^{-6} | 17,761 | 17,146 | 18,350 | 18,163 | 19,020 | 18,030 | 18,785 | 18,136 | 18,606 | 18,235 |
| 10 | 665 | 2.261×10^{-6} | 23,792 | 23,314 | 26,379 | 26,464 | 26,530 | 27,031 | 24,654 | 26,131 | 25,747 | 25,143 |
| 15 | 830 | 1.452×10^{-6} | 35,885 | 36,770 | 36,844 | 37,163 | 39,360 | 39,244 | 40,608 | 38,193 | 38,338 | 37,903 |
| 20 | 1120 | 0.797×10^{-6} | 47,305 | 49,034 | 52,173 | 49,448 | 52,162 | 50,740 | 52,090 | 51,832 | 51,053 | 51,097 |

TABLE 15.7
Regression Results

| Run | $\hat{\beta}_0$ | $\hat{\beta}_1$ | Residual error
variance estimate
s^2 |
|----------|-----------------|-----------------|--|
| 1 | 1111.53 | 2328.62 | 0.884456 |
| 2 | 1680.40 | 2407.08 | 1.427593 |
| 3 | 1698.96 | 2491.77 | 0.755366 |
| 4 | 950.91 | 2527.44 | 1.499050 |
| 5 | 1351.76 | 2453.38 | 0.349909 |
| Combined | 1358.71 | 2441.66 | 2.032306 |

TABLE 15.8
Analysis of Variance Table for Differences Between Runs

| Source | df | SS | MS | F |
|--------------------------------|----|-------------|-------------|-------|
| Common regression | 1 | 28,549.3405 | 28,549.3405 | — |
| Between intercepts | 4 | 45.7356 | 11.4339 | 11.63 |
| Between slopes | 4 | 22.9745 | 5.7436 | 5.84 |
| Total regression | 9 | 28,618.0506 | — | — |
| Lack of fit | 20 | 25.8375 | 1.2919 | 1.66 |
| Between replicate measurements | 30 | 23.3262 | 0.7775 | — |
| Residual from regression | 50 | 49.1637 | 0.9833 | — |
| Total sum of squares | 59 | 28,667.2143 | | |

mon regression” and the differences “between intercepts.” Because the linear forms determining the slopes and the intercepts are not orthogonal, at least one of the sums of squares constituting the total amount explained by the individual regressions must be determined by difference.

5. The sum of squares between replicate measurements is determined from the weighted sum of squares of the 30 differences between the two measurements on each standard.

6. The lack of fit sum of squares is the difference between the “residual from regression” and the “between replicate measurements” sums of squares.

If the weights correctly reflect the reciprocals of the variances of the individual measurements, then the mean square between replicate measurements would be expected to be 1.0. The observed value of 0.7776 is actually less than this expected value, although not significantly, so there is no reason to suspect unexpected measurement variability or the validity

of the weights used. The F-ratio between the "lack of fit" mean square, which represents the deviations of the average of the replicate measurements from the individually fitted lines, and the mean square "between replicate measurements" is also not significant, indicating the adequacy of the linear calibration function. This is supported by the fact that the mean square for the combined "residual from regression" has almost exactly the expected value of 1.00. The remaining tests are performed using this residual mean square in the denominator of the calculated F-ratios because the lack-of-fit test was not significant.

There is no point in this example in testing for the significance of the "common regression" because the purpose of the calibration exercise is to determine an assumed functional relationship. The question of interest is whether or not there is evidence that a single calibration equation is adequate for the combined data. Both the "between slopes" and "between intercepts" F-ratios are significant at the 0.01 level, so there is evidence that neither the slope nor the intercept remains constant from one calibration run to another. It is necessary to understand how the changes occur before a good decision can be made concerning which calibration parameter estimates to use when measuring waste barrels from the process during the current time period. Two cases are presented here to illustrate some of the considerations involved in the selection of appropriate estimates for the calibration parameters.

For the first case, suppose that sufficient calibration and calibration-check data and information have been accumulated over recent time periods to demonstrate that the calibration is stable within each period and that the changes occur between periods. An example of a system that might behave in such a manner would be a stable measurement system which has routine maintenance performed at the end of each month. This routine maintenance might include replacement of some of the components with new or reconditioned components. For measurement systems subject to these types of changes, the calibration parameter estimates obtained from the calibration run made during the current time period are appropriate for use throughout that period. Because Run 5 is from the current time period in this example, $\hat{\beta}_0 = 1351.76$ and $\hat{\beta}_1 = 2453.38$ are the parameter estimates to use when measuring waste barrels from the process during the fifth time period. To estimate the ^{235}U content, the usual relationship

$$\hat{x}_p = \frac{y_p - \hat{\beta}_0}{\hat{\beta}_1}$$

would be used with these values for $\hat{\beta}_0$ and $\hat{\beta}_1$.

There is some question concerning which numerical values to use for the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ and the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$ when calculating the variance of \hat{x} due to calibration. All five calibration runs tend to confirm that the weights for the weighted regression analysis were correctly assigned. Thus, the assigned value of $\sigma^2 = 1.0$ should produce better values for the needed variances and covariance than the sample estimates obtained from any of the individual run estimates. To illustrate the calculations for this case, suppose that two waste barrels were counted for 100 s each during the fifth time period and that 13,919 and 42,267 were the recorded net counts. The sum of the ^{235}U in inventory in the two waste barrels is needed along with the random error variance for the sum and the variance for the sum due to the uncertainty in the calibration parameter estimates. The same historical information that provided the σ_i values used for assigning weights to the calibration data indicate that the random measurement standard deviation for the first waste barrel should be approximately 508 and for the second barrel approximately 928. Following the same calculation procedures as in Example 15.3 (Section 15.1.2.2) the estimate for the desired sum is

$$\begin{aligned}\hat{x}_{\text{sum}} &= \frac{(y_{p1} + y_{p2}) - 2\hat{\beta}_0}{\hat{\beta}_1} \\ &= \frac{(13,919 + 42,267) - 2(1351.76)}{2453.38} = 21.80 \text{ g}\end{aligned}$$

and the component of the estimated variance due to measurement error is

$$\text{Vâr}_r(\hat{x}_{\text{sum}}) \cong \frac{(508)^2 + (928)^2}{(2453.38)^2} = 0.1860 \text{ (g)}^2$$

The component of the variance in the estimated sum due to uncertainty in the calibration relationship is

$$\begin{aligned}\text{Var}_c(\hat{x}_{\text{sum}}) &\cong \left(\frac{-2}{\hat{\beta}_1} \right)^2 \text{Var}(\hat{\beta}_0) \\ &+ \left(-\frac{y_{p1} + y_{p2} - 2\hat{\beta}_0}{\hat{\beta}_1^2} \right)^2 \text{Var}(\hat{\beta}_1) \\ &+ 2 \left(\frac{-2}{\hat{\beta}_1} \right) \left(-\frac{y_{p1} + y_{p2} - 2\hat{\beta}_0}{\hat{\beta}_1^2} \right) \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)\end{aligned}$$

As discussed previously, if it is assumed that $\sigma^2 = 1$ the variances and covariances depend only on the standard values and the weights common to all runs, and are the same for each run. For an individual run involving five duplicate responses, a computation parallel to that given in Example 15.2 produces the values

$$\begin{aligned} \hat{\text{V}}\text{ar}(\hat{\beta}_0) &= 64,187.70 \\ \hat{\text{V}}\text{ar}(\hat{\beta}_1) &= 1044.10 \\ \hat{\text{C}}\text{ov}(\hat{\beta}_0, \hat{\beta}_1) &= -6167.78 \end{aligned}$$

so the component of the estimated variance due to calibration is

$$\begin{aligned} \hat{\text{V}}\text{ar}_c(\hat{x}_{\text{sum}}) &\cong \left[-\frac{2}{2453.38} \right]^2 (64,187.70) \\ &\quad + \frac{(53,482.5)^2}{(2453.38)^4} (1044.10) \\ &\quad + \frac{4(53,482.5)}{(2453.38)^3} (-6167.78) \\ &= 0.04266 + 0.08243 - 0.08935 \\ &= 0.03574 \text{ (g)}^2 \end{aligned}$$

A second type of changing calibration is quite different from the previous case. For this second case, it is assumed that at any time the system is recalibrated the parameter estimates from the repeated calibrations will change randomly, and that the magnitude of these shifts can be described by the variation observed in parameter estimates from the available calibration runs. With these assumptions, an average calibration curve (in some sense) could be used, supported by the additional assumption that there is a true calibration curve which is fixed rather than variable, so that the observed variability in the parameter estimates is due to the calibration process. The changing calibration curves observed from the various calibration runs might be the result of random short-term fluctuations in environmental conditions which affect the instrument response. Such factors as day-to-day changes in the line voltage to an NDA instrument might cause such behavior. The purpose of recalibrating such a system is to obtain a better estimate of the average calibration curve and the variances and covariances of the parameter estimates. It is generally recommended that the repeated calibrations be spread out over time to permit monitoring for nonrandom shifts or trends, should they occur.

Obtaining the average, sample variances, and sample covariance of the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, is straightforward when all of the repeated calibration runs have the same design and are performed identically. If different numbers of replicate measurements on the standards or different choices of standards have occurred for some of the repeated calibration runs, then an alternate weighted analysis approach to the one described below must be used. The individual estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ for each run and for the combined data were given in Table 15.7. The average value of the individual estimates is, in this case of repeated runs with the same design, identical to the estimates obtained from the combined data. The estimated variances and covariance obtained from the five pairs of individual estimates are

$$\begin{aligned}\hat{\text{Vâr}}(\hat{\beta}_0) &= 111,675.40 \\ \hat{\text{Vâr}}(\hat{\beta}_1) &= 5995.10 \\ \hat{\text{Côv}}(\hat{\beta}_0, \hat{\beta}_1) &= -298.86\end{aligned}$$

For this case, the sum of the ^{235}U in the two waste barrels would be calculated from the average (or, equivalently, combined) estimates as

$$\begin{aligned}\hat{x}_{\text{sum}} &= \frac{(13,919 + 42,267) - 2(1358.71)}{2441.66} \\ &= \frac{53,468.58}{2441.66} = 21.90 \text{ g}\end{aligned}$$

The component of the estimated variance due to measurement error would be estimated as

$$\hat{\text{Vâr}}_r(\hat{x}_{\text{sum}}) \cong \frac{(508)^2 + (928)^2}{(2441.66)^2} = 0.1877 (\text{g})^2$$

and the component of the estimated variance due to uncertainty in the calibration relationship as

$$\begin{aligned}\hat{\text{Vâr}}_c(\hat{x}_{\text{sum}}) &\cong \frac{4}{(2441.66)^2} (111,675.40) \\ &\quad + \frac{(53,468.6)^2}{(2441.66)^4} (5995.10) \\ &\quad + \frac{4(53,468.6)}{(2441.66)^3} (-298.86) \\ &= 0.07493 + 0.48223 - 0.00439 \\ &= 0.55277 (\text{g})^2\end{aligned}$$

The estimates of the sum for the two cases agree within 0.1 g in 22 g, or about 0.5%, and the measurement variances are essentially the same. However, the variance component due to calibration uncertainties is considerably larger in the second case, as would be expected because of the inclusion of the variability between runs. There may be situations where the two estimates for the sum would be in poorer agreement. As an example, if Run 4 had been the most recent, the estimate for the sum in the first case would have been 21.48 g as compared with an estimate in the second case of 21.90 g. It is important to understand how changes are occurring when a calibration curve is not constant from one run to another, and to choose the most realistic estimates for the calibration parameters.

15.1.3 Nonlinear Calibration

Many of the instrument and vessel calibration applications in safeguards can be handled adequately using methods of linear regression analysis. For some NDA instrument applications, however, the curve may depart from linearity, depending on the range of standards covered by the calibration. In the case of process vessel calibrations, the curve may be nonlinear in a complex way, depending on the shape of the vessel and the possible presence of internal piping or mixing equipment. In some cases, the nonlinear calibration curve can be adequately represented using a quadratic or possibly a cubic polynomial model, providing that the range of the calibration can be restricted sufficiently or that the total range can be broken into subranges to limit the curvature within individual calibration intervals. It is also possible that a simple transformation of one or both of the variables in a nonlinear relationship will result in a linear relationship between the transformed variables. The example used in the following section was selected so that both the use of a quadratic calibration model and the use of transformations can be demonstrated on the same data set.

Small data sets are used in this chapter to illustrate various estimation problems. The use of such sets should not be interpreted as a recommendation for sample sizes or numbers of standards to be used in any particular application. Replicate measurements may be desirable on some or all of the standards, and more than a single calibration run may be performed, as considered in Section 5.2.3.2.

15.1.3.1 Simple Nonlinear Calibration

The fitting of simple nonlinear calibration models will be illustrated by the following example.

Example 15.5 A high-level neutron coincidence counter (HLNCC) is used to measure the grams of ^{240}Pu effective in samples of PuO_2 based on neutron coincidence counting (Krick, 1980, Appendix B). Six standards were used to calibrate the instrument and 1000 s counts were obtained on each standard. The data from single measurements on each standard are given in the following table and are shown in Figure 15.3. These measurements differ from those in Example 15.3 in that the count rates are not corrected for neutron multiplication.

| ^{240}Pu effective
(g) | Real coincidence
response
(counts/s) |
|------------------------------------|--|
| 97.45 | 715.4 |
| 195.0 | 1609.4 |
| 292.3 | 2577.0 |
| 389.8 | 3553.6 |
| 487.4 | 4617.4 |
| 584.3 | 5811.8 |

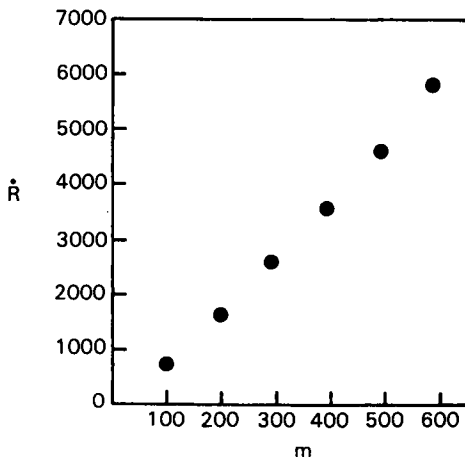


Figure 15.3 Coincidence counting results on six calibration standards.

Also, an unknown sample was counted for which the measured response was 2791.0 counts/s. From other analytical measurements on the unknown sample, it had an assigned value of 312.6 g of ^{240}Pu effective.

First, a quadratic model will be used to fit these calibration data, assuming

$$\dot{R} = \beta_0 + \beta_1 m + \beta_2 m^2 \quad (15.9)$$

where

\dot{R} = real coincidence response

m = ^{240}Pu effective

Using standard computational procedures, the following estimates were obtained for the quadratic model:

$$\hat{\beta}_0 = -84.00115299$$

$$\hat{\beta}_1 = 7.958176756$$

$$\hat{\beta}_2 = 0.0035933502$$

$$\text{Vâr}(\hat{\beta}_0) = 2499.577872$$

$$\text{Vâr}(\hat{\beta}_1) = 0.112780025$$

$$\text{Vâr}(\hat{\beta}_2) = 2.324774213 \times 10^{-7}$$

$$\text{Côv}(\hat{\beta}_0, \hat{\beta}_1) = -15.64001796$$

$$\text{Côv}(\hat{\beta}_0, \hat{\beta}_2) = 2.058836758 \times 10^{-2}$$

$$\text{Côv}(\hat{\beta}_1, \hat{\beta}_2) = -1.585145973 \times 10^{-4}$$

The calibration equation can be rewritten as

$$\beta_2 m^2 + \beta_1 m + (\beta_0 - \dot{R}) = 0 \quad (15.10)$$

The positive solution for m is

$$m = \frac{-\beta_1 + \sqrt{\beta_1^2 - 4\beta_2(\beta_0 - \dot{R})}}{2\beta_2}$$

The estimated values $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ were used, along with the observed \dot{R} value of 2791.0, to obtain an estimate \hat{m} of 316.1368 g for the content of the unknown sample. As in previous examples, Equation 3.51 can be used to calculate the approximate variance of \hat{m} due to the uncertainty in the estimated calibration parameters. There is also uncertainty in the predicted value \hat{m} due to the random error in measuring \dot{R} . This uncertainty will not

be included in the following calculation but will be considered later when a method of transforming the calibration data is illustrated.

To obtain an estimate for the variance of \hat{m} due to calibration uncertainty using the Taylor's series approach, the following partial derivatives, given in implicit form, are needed:

$$\frac{\partial m}{\partial \beta_0} = -\frac{1}{2\beta_2 m + \beta_1} \quad (15.11)$$

$$\frac{\partial m}{\partial \beta_1} = -\frac{m}{2\beta_2 m + \beta_1} \quad (15.12)$$

$$\frac{\partial m}{\partial \beta_2} = -\frac{m^2}{2\beta_2 m + \beta_1} \quad (15.13)$$

From Equation 3.51,

$$\begin{aligned} \text{Vâr}_c(\hat{m}) &\cong \left(\frac{\partial m}{\partial \beta_0} \right)^2 \text{Vâr}(\hat{\beta}_0) + \left(\frac{\partial m}{\partial \beta_1} \right)^2 \text{Vâr}(\hat{\beta}_1) \\ &+ \left(\frac{\partial m}{\partial \beta_2} \right)^2 \text{Vâr}(\hat{\beta}_2) \\ &+ 2 \left(\frac{\partial m}{\partial \beta_0} \right) \left(\frac{\partial m}{\partial \beta_1} \right) \text{Côv}(\hat{\beta}_0, \hat{\beta}_1) \\ &+ 2 \left(\frac{\partial m}{\partial \beta_0} \right) \left(\frac{\partial m}{\partial \beta_2} \right) \text{Côv}(\hat{\beta}_0, \hat{\beta}_2) \\ &+ 2 \left(\frac{\partial m}{\partial \beta_1} \right) \left(\frac{\partial m}{\partial \beta_2} \right) \text{Côv}(\hat{\beta}_1, \hat{\beta}_2) \end{aligned} \quad (15.14)$$

where the derivatives are evaluated at the estimates of m , β_1 , and β_2 . From the estimate \hat{m} obtained using Equation 15.10 and the estimates $\hat{\beta}_2$ and $\hat{\beta}_1$ previously calculated, estimates of the derivatives can be calculated from Equations 15.11, 15.12, and 15.13 as -0.0977502 , -30.902440 , and -9769.3997 , respectively. Substituting these values along with the estimated variances and covariances of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ into Equation 15.14, the estimated variance of the estimate \hat{m} is obtained as

$$\begin{aligned} \text{Vâr}_c(\hat{m}) &= 23.8837 + 107.7005 + 22.1879 - 94.4882 \\ &\quad + 39.3222 - 95.7106 = 2.8955 \text{ (g)}^2 \end{aligned}$$

$$\text{SÊ}_c(\hat{m}) = 1.7016 \text{ g}$$

Direct solution of the calibration equation for m is only practical for linear and quadratic equations when polynomial models are used for non-linear calibrations. Numerical solutions of the calibration equation, along with numerical estimates of the variances of these estimates calculated using the implicit form of the derivatives, are advised for higher order polynomials.

Experience has shown that an appropriate transformation of one or both variables may result in a more nearly linear relationship between the transformed variables. A second possible benefit from making a transformation on the response variable is that it may result in increased homogeneity in the variances of the observed responses. For NDA calibration data similar to the HLNCC data used in Example 15.5, both of these objectives can be accomplished by taking natural logarithms of both variables. A linear calibration model in the transformed variables can be written as

$$\ln \hat{R} = \beta_0 + \beta_1 \ln m \quad (15.15)$$

This is equivalent to assuming the power function relationship

$$\hat{R} = \beta'_0 m^{\beta_1} \quad (15.16)$$

where $\beta'_0 = e^{\beta_0}$, between the untransformed variables. The calibration data are plotted on a logarithmic scale in Figure 15.4, and it appears that the transformed variables are very nearly linearly related. In this example there should be some concern about whether the observed responses—which range from about 700 to 5800—should be expected to have constant variance. With no replication of measurements for the six standards, it is not possible to really investigate this question nor to demonstrate stabilization of the proposed variance through use of the transformation. However, experience with other NDA calibration data sets similar to these HLNCC data has indicated that a logarithmic transformation of the response variable often helps to increase the homogeneity of the variances.

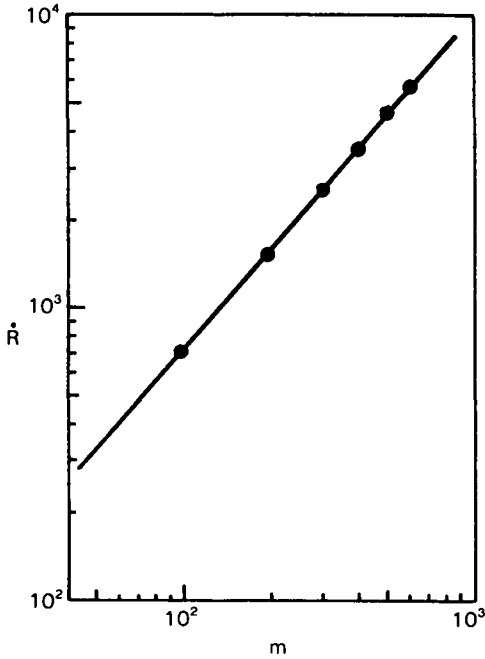


Figure 15.4 Transformed calibration data.

From the computational procedures for unweighted linear regression, using the logarithms of the standard values and coincidence responses, the following estimates were obtained:

$$\hat{\beta}_0 = 1.247461137$$

$$\hat{\beta}_1 = 1.163022395$$

$$\text{Var}(\hat{\beta}_0) = 0.0011608119$$

$$\text{Var}(\hat{\beta}_1) = 0.0000356282$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -0.000202221$$

Solving Equation 15.15 for $\ln m$ yields

$$\ln m = \frac{\ln \dot{R} - \beta_0}{\beta_1}$$

or

$$m = \exp\left(\frac{\ln \dot{R} - \beta_0}{\beta_1}\right)$$

For the container of plant material with $\dot{R} = 2791.0$, the estimate of the ^{240}Pu effective is $\hat{m} = 314.0057497$ g. If the two-parameter form of Equation 15.14 is used with

$$\frac{\partial m}{\partial \beta_0} = \left(\frac{-1}{\beta_1}\right) \exp\left(\frac{\ln \dot{R} - \beta_0}{\beta_1}\right) \quad (15.17)$$

$$\frac{\partial m}{\partial \beta_1} = \left(-\frac{\ln \dot{R} - \beta_0}{\beta_1^2}\right) \exp\left(\frac{\ln \dot{R} - \beta_0}{\beta_1}\right) \quad (15.18)$$

the estimated variance and standard error of this estimate are

$$\begin{aligned} \text{Vâr}_c(\hat{m}) &= (-269.9911464)^2(0.0011608119) \\ &\quad + (-1552.290147)^2(0.0000356282) \\ &\quad + 2(-269.9911)(-1552.2901)(-0.000202221) \\ &= 84.617608 + 85.849873 - 169.503467 \\ &= 0.9640 \text{ (g)}^2 \end{aligned}$$

and

$$\hat{\text{SĒ}}_c(\hat{m}) = 0.9818 \text{ g}$$

In this simple calibration example, the model using the transformed data appears to fit the data somewhat better than the quadratic and has a smaller variance due to calibration uncertainty for containers in the range of 314 g of ^{240}Pu effective.

In addition to the variance due to calibration uncertainty, there is also a variance due to the random error of the response variable, \dot{R} , when a container is measured using the instrument. Because there were no replicate measurements on any of the standards during the calibration run, there is no direct measure of the variability of the measurement. However, because a fixed counting time was used in all of the standards measurements, and assuming that the logarithmic transformation did result in stabilization of the variance, then the residual variance estimate from the

regression analysis provides an estimate of the variance of $\ln \hat{m}$. This contribution to the variance of a predicted \hat{m} is

$$\text{Var}_r(\hat{m}) \cong \left(\frac{\partial m}{\partial \ln \hat{R}} \right)^2 \text{Var}(\ln \hat{R}) \quad (15.19)$$

where the partial derivative with respect to $\ln \hat{R}$ is equal to the negative of the partial derivative with respect to β_0 , and the residual error variance from the log-linear fit is used as an estimate of $\text{Var}(\ln \hat{R})$. For the unknown container, the estimate of this component of the variance turns out to be

$$\text{Vâr}_r(\hat{m}) \cong (269.9911464)^2(0.0000781836) = 5.6992$$

This component should be added to the component due to calibration as in Example 15.4.

15.1.3.2 Multiple Data Sets

As noted in Section 15.1.2.3, all measurement systems need to be monitored routinely for shifts in the calibration from one time period to another. The monitoring over time of the calibration should be accomplished by routinely making at least one calibration check during each time period. If the calibration shifts occur randomly over time, the calibration parameter estimates may be regarded as being randomly selected from a population of parameters and the variability of the estimates can then be more realistically measured from several calibrations obtained over time than from a single calibration run (or from several calibration runs made very close together in time). When the calibration shifts do not occur randomly over time, decision rules need to be established either for bias correction of the measurement results or for complete recalibration.

In this section, it is assumed that a quadratic model is adequate to represent the nonlinear calibration curve over the range covered by the standards used for calibration. The adequacy of this assumption has been demonstrated in a number of applications, especially for NDA calibration data sets. It is further assumed that there is a constant variance over the range of calibration. The methods are demonstrated through an example involving two similar sets of data. The first set is based on actual repeated calibration runs made using an NDA instrument known as a stabilized assay meter (SAM). Because an analysis of the actual calibration data did not indicate a difference in the calibration curves among the runs and

there were no replicate measurements made on the standards during the individual calibration runs, a second set of calibration data was generated to demonstrate a more in-depth analysis than was possible using the original data. The synthesized data for the second example assumed the same general model used in the analysis of the original data, but included randomly imposed run-to-run differences in the calibration parameters. Because experience with NDA calibration data has indicated that it is difficult to maintain constant variance (or constant relative variance) when there is more than a factor of 10 difference between the assigned value of the high standard and the assigned value of the low standard, the range of standards used for the data in the second set was restricted to between 3.9761 and 40.18 g plutonium. It is frequently difficult to maintain constant variance over the range of the standards, and when the range of the standards used is too great, it may not be possible to find a simple calibration model to adequately describe the data.

Example 15.6 A stabilized assay meter (SAM) was calibrated using six standards. Five different calibration runs were made within a period of 2 days. The following calibration data were obtained:

TABLE 15.9
SAM Calibration Data

| Assigned value
of the standard
(g Pu) | Net counts | | | | |
|---|------------|---------|---------|---------|---------|
| | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 |
| 1.0019 | 653 | 680 | 577 | 606 | 687 |
| 1.9971 | 1293 | 1343 | 1362 | 1280 | 1318 |
| 3.9761 | 2779 | 2690 | 2704 | 2767 | 2792 |
| 10.571 | 6937 | 6822 | 6945 | 6909 | 6902 |
| 40.18 | 27,698 | 27,666 | 27,574 | 27,849 | 27,679 |
| 160.15 | 101,511 | 101,810 | 101,941 | 102,163 | 101,602 |

A model adequate to describe these calibration data is

$$\ln C = \beta_0 + \beta_1 (\ln g) + \beta_2 (\ln g)^2$$

where

C = net counts

g = grams plutonium contained in the standards

The following estimates for the three calibration parameters (β_0 , β_1 , β_2) were found:

| Run | β_0 | β_1 | β_2 | Residual variance |
|----------|-----------|-----------|------------|-------------------|
| 1 | 6.4713 | 1.0438 | -0.0090647 | 0.001355 |
| 2 | 6.5134 | 1.0018 | -0.0020567 | 0.001146 |
| 3 | 6.4001 | 1.0994 | -0.0176234 | 0.002971 |
| 4 | 6.4139 | 1.0842 | -0.0147685 | 0.001523 |
| 5 | 6.5161 | 1.0114 | -0.0042405 | 0.001603 |
| Combined | 6.4630 | 1.0481 | -0.0095508 | 0.001577 |

To make a test for differences among the five calibration runs, the analysis of variance shown in Table 15.10 was calculated:

TABLE 15.10

Analysis of Variance Table for Five Calibration Runs*

| Source | df | SS | MS |
|------------------------------------|----|---------------------------------|---------------------|
| Between runs | 12 | $SS_1 = SS_2 - SS_3 = 0.016785$ | $MS_1 = 0.00139875$ |
| Residual (combined data) | 27 | $SS_2 = 0.042579$ | $MS_2 = 0.001577$ |
| Residual (5 individual run models) | 15 | $SS_3 = 0.025794$ | $MS_3 = 0.0017196$ |

*Transformed data using logarithms.

In this table, SS_2 is the residual sum of squares and MS_2 is the residual variance estimate when a single calibration curve is fitted to the combined data. $SS_3 = 3(0.001355 + \dots + 0.001603)$ is the total residual sum of squares from fitting calibration curves to the individual run data. Because the mean square "between runs" (due to fitting 12 additional parameters in the individual calibration models) is not significantly larger ($F < 1$) than the residual variance estimate from the individual runs, there is no evidence of significant differences among the calibration runs. For the time period immediately following the five calibration runs, the best estimates for the calibration parameters are those obtained from the analysis of the combined data (from all five calibration runs). From the analysis of the combined data, the following variance and covariance estimates were obtained using the usual calculational procedures:

$$\hat{\text{Vâr}}(\hat{\beta}_0) = 0.01504$$

$$\hat{\text{Vâr}}(\hat{\beta}_1) = 0.01523$$

$$\hat{\text{Vâr}}(\hat{\beta}_2) = 0.002883$$

$$\hat{\text{Côv}}(\beta_0, \hat{\beta}_1) = -0.0001771$$

$$\hat{\text{Côv}}(\beta_0, \hat{\beta}_2) = +0.00002737$$

$$\hat{\text{Côv}}(\hat{\beta}_1, \hat{\beta}_2) = -0.00004225$$

These variances and covariances can be used to calculate an estimate of the standard deviation of the computed plutonium content for an actual production item using methods similar to those used in Example 15.5 of Section 15.1.3.1.

As with all measurement systems, the calibration for this measuring instrument should be monitored during subsequent time periods following the original calibration runs. If the monitoring consists of one new calibration run, an analysis of variance test analogous to the one used to test for differences among the original runs can be used to check for a change in the calibration equation.

Because there were no replicate measurements made on any of the standards during any of the individual calibration runs, there was no estimate of the pure error variance available from the original set of calibration data. Therefore, a good test for lack of fit of the model was not available. There is a strong indication, however, that the model used for the calibration is not completely adequate to describe the observed calibration data. The predicted net counts for the 10.571, 40.18, and 160.15 standards are 7,197, 27,004, and 102,455, respectively, based on the parameter estimates from the analysis of the combined data. For these standards, the predicted values are completely outside the range of observed counts from the five calibration runs, an indication that the model is not adequate to cover the entire range of standards used for the calibration. This observation illustrates the difficulty of finding a simple calibration model to adequately describe the data when assigned values of the standards differ by more than a factor of 10.

A second set of data will permit a demonstration of a lack-of-fit test, in addition to a test for differences among several calibration runs. The data were synthesized from the assumed model and the parameter values obtained from an analysis of the observed calibration results for the 3.9761, 10.571, and 40.18 standards. Run-to-run differences in the parameters were randomly imposed and replicate measurements were generated for each standard during each calibration run. For reasons discussed earlier, the range of standards used was restricted, the lowest standard containing 3.9761 g of plutonium and the highest standard 40.18 g of plutonium. The synthesized data are presented in Table 15.11.

TABLE 15.11
Synthesized SAM Calibration Data

| Assigned value
of the standards
(g Pu) | Net counts | | | | | | | | | |
|--|------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | Run 1 | | Run 2 | | Run 3 | | Run 4 | | Run 5 | |
| 3.9761 | 2731 | 2680 | 2678 | 2693 | 2743 | 2736 | 2736 | 2758 | 2769 | 2749 |
| 6.580 | 4336 | 4292 | 4410 | 4482 | 4348 | 4381 | 4328 | 4326 | 4379 | 4387 |
| 10.571 | 6954 | 6762 | 7019 | 7110 | 6990 | 6793 | 6745 | 6781 | 6780 | 6828 |
| 20.25 | 12,987 | 13,082 | 14,016 | 13,755 | 13,231 | 13,207 | 13,111 | 13,205 | 13,032 | 13,202 |
| 40.18 | 27,605 | 27,272 | 28,952 | 28,434 | 27,403 | 27,273 | 27,792 | 27,601 | 27,263 | 27,458 |

Using the same calibration model as in the first example, the estimates obtained for the parameters are given in the following table:

| Run | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | Residual variance |
|----------|-----------------|-----------------|-----------------|-------------------|
| 1 | 6.7443 | 0.78147 | 0.042968 | 0.00014179 |
| 2 | 6.6028 | 0.90695 | 0.022845 | 0.00010718 |
| 3 | 6.7429 | 0.79585 | 0.039099 | 0.00006562 |
| 4 | 6.8249 | 0.71503 | 0.056005 | 0.00001906 |
| 5 | 6.8100 | 0.73976 | 0.049405 | 0.00004142 |
| Combined | 6.7450 | 0.78781 | 0.042064 | 0.00032055 |

The analysis of variance table for these calibration runs is presented in Table 15.12. The analysis is quite similar to that of Example 15.4.

TABLE 15.12

Analysis of Variance Table for Five Calibration Runs†

| Source | df | SS | MS |
|---|-------------|------------|-------------------------------|
| Total regressions | 15 - 1 = 14 | 33.4689404 | |
| Common regression
(combined data) | 3 - 1 = 2 | 33.4565001 | MS ₁ = 16.72825005 |
| Reduction due to fitting
additional parameters in
individual models | 12 | 0.0124403 | MS ₂ = 0.00103669 |
| Residual from regression | 35 | 0.0026255 | MS ₃ = 0.00007501 |
| Lack of fit | 10 | 0.0006230 | MS ₄ = 0.00006230 |
| Between replicate
measurements | 25 | 0.0020025 | MS ₅ = 0.00008010 |
| Total sum of squares | 50 - 1 = 49 | 33.4715659 | |

†Transformed using logarithms—synthesized example.

All entries in the table are calculated from the logarithms of the observations. Beginning at the bottom of Table 15.12, the entries are calculated as follows:

1. The **total sum of squares** (33.4715659) is the sum of squares of the deviations from the mean for the combined data. The degrees of freedom (df) are the total number of observations minus one.

2. The sum of squares **between replicate measurements** (0.0020025) was obtained by accumulating the sum of squared deviations of the replicate measurements for each set of measurements from the mean for that set. Because only two replicate measurements were routinely made in this example, this quantity can be calculated by squaring the difference between the paired results, accumulating over all standards and runs, and then dividing by 2. The df in this case are one for each pair of replicate measurements. More generally, the number is the total number of observations less one df for each set.

3. The **lack of fit** sum of squares (0.0006230) is obtained as the difference between the "residual from regression" sum of squares and the "between replicate measurements" sum of squares. The df are also obtained by difference.

4. The **residual from regression** sum of squares (0.0026255) is the sum of the individual residual sums of squares from the analysis of each run. The residual for each run has $10 - 3 = 7$ df and the total has $5(7) = 35$ df.

5. The **reduction due to fitting additional parameters in individual models** sum of squares (0.0124403), which is due to fitting 12 additional parameters after having fitted 3 parameters to the combined data, is again obtained by difference. The total residual sum of squares from regression and the sum of squares explained by the common regression are subtracted from the total sum of squares. The df are the additional number of parameters estimated in the individual run calibration equations, and can also be obtained by difference ($49 - 35 - 2 = 12$).

6. The **common regression** sum of squares (33.4565001) is calculated directly from the combined data as described in Section 6.4.1, Equation 6.18. As shown in Table 6.2, it is the difference between the residual from the common regression and the total sum of squares, and has $p - 1 = 3 - 1 = 2$ df.

As in Example 15.4, all entries are obtained from the regression analysis of variance on the individual runs and the combined runs, except the sum of squares between replicates. The mean squares are obtained by dividing the sums of squares by the corresponding degrees of freedom.

To test for lack of fit, form the F-ratio MS_4/MS_5 . In this example, the calculated F is less than one, and there is no evidence that the chosen model is inadequate for the calibration data being analyzed. In such a situation, MS_3 is used as an estimate of the random error variance. The F-ratio $16.73/.000045$ for the common regression is very large, as would be expected. To test whether there is a significant difference among the

individual calibration equations, the F-ratio MS_2/MS_3 is used. The value calculated for this F-ratio is 13.82. Because the 99% fractile for the F-distribution with numerator df of 12 and denominator df of 35 is equal to 2.74, the hypothesis of no difference among the calibration runs is rejected, so there is evidence of differences between runs.

This second analysis demonstrates that with the added calibration data from making replicate measurements on the standards, it is possible to test for the adequacy of the chosen calibration model in addition to testing for differences due to the different calibration runs. The ability to make both types of test is important in many practical applications.

15.1.4 The Cumulative Error Model

The cumulative error model was introduced by Mandel (1957) to provide the correct treatment of a class of calibration problems for which the usual least squares model is not appropriate. In the example which gave rise to the model, a plastic specimen is tested for its resistance to abrasion by subjecting it to the actions of an abrasive wheel for successive test periods. The lengths of the test periods are measured without error. The model used for the amount of wear is

$$y_i = \beta x_i + e_i \quad (15.20)$$

where x_i is the total time of exposure prior to the i^{th} measurement of wear, y_i is the amount of wear, and e_i is the error in y_i . After an exposure of x_1 units of time, the wear is y_1 ; after x_2 units of time, the wear is y_2 , where $x_2 > x_1$, $y_2 \geq y_1$, and so on. The error in y_i is cumulative and arises from the variation in the amount of wear per unit of exposure time and not from the variation in measuring the amount of wear. This model has been particularly useful in material accounting in connection with the calibration of process vessels, where the dominating source of variation is often the cumulative error in the measurement of the incremental additions of liquid, not the instrumental response (Jaech, 1964).

Assume that the data from the study arise from i incremental additions starting with $x_0 = y_0 = 0$ resulting in the observation (x_i, y_i) . For the particular conditions considered here, both the measured response and the variance of the measured response after any incremental addition are assumed to be proportional to the total added. This means that

$$y_i - y_{i-1} = \beta(x_i - x_{i-1}) + \epsilon_i \quad (15.21)$$

where the $\epsilon_i = e_i - e_{i-1}$, $i = 1, \dots, n$, are independent, and that

$$\text{Var}(y_i - y_{i-1}) = \text{Var } \epsilon_i = (x_i - x_{i-1})\sigma^2 \quad (15.22)$$

where σ^2 is the incremental variance per unit of incremental addition. The expression *cumulative error model* arises from the fact that the errors e_i in Equation 15.20 are the cumulative sums

$$\sum_{j=1}^i \epsilon_j$$

of the independent errors in the increments.

From Equation 15.22 it follows that the quantities $\eta_i = \epsilon_i / \sqrt{x_i - x_{i-1}}$ would not only be independent but would have constant variance σ^2 . Dividing Equation 15.21 on both sides by $\sqrt{x_i - x_{i-1}}$ gives

$$z_i = \beta t_i + \eta_i \quad (15.23)$$

where

$$z_i = \frac{y_i - y_{i-1}}{\sqrt{x_i - x_{i-1}}}$$

$$t_i = \sqrt{x_i - x_{i-1}}$$

The transformed model $z_i = \beta t_i + \eta_i$ can be fitted by the ordinary least squares method of Chapter 6 because the z_i have equal variances and are uncorrelated. The estimated slope of a first degree model that passes through the origin is

$$\hat{\beta} = \frac{\sum z_i t_i}{\sum t_i^2} = \frac{\sum (y_i - y_{i-1})}{\sum (x_i - x_{i-1})} = \frac{y_n}{x_n} \quad (15.24)$$

since $y_0 = x_0 = 0$. The variance of the estimate $\hat{\beta}$ is

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum t_i^2} = \frac{\sigma^2}{x_n} \quad (15.25)$$

From ordinary least squares theory it is known that the residual variance is estimated by

$$s^2 = \frac{\sum z_i^2 - \hat{\beta} \sum t_i z_i}{n-1} = \frac{\sum \frac{(y_i - y_{i-1})^2}{x_i - x_{i-1}} - \frac{y_n^2}{x_n}}{n-1}$$

and thus the variance of $\hat{\beta}$ is estimated by

$$\text{Vâr}(\hat{\beta}) = \frac{s^2}{\sum t_i^2} = \frac{s^2}{x_n} \quad (15.26)$$

When the model is of the form

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i \\ &= \beta_0 + \beta_1 x_i + \sum_1^i \epsilon_j \end{aligned}$$

it can be shown that

$$\begin{aligned} \hat{\beta}_1 &= \frac{y_n - y_1}{x_n - x_1} \\ \hat{\beta}_0 &= y_1 - \beta_1 x_1 \\ s^2 &= \frac{\sum \frac{(y_i - y_{i-1})^2}{x_i - x_{i-1}} - \frac{(y_n - y_1)^2}{x_n - x_1}}{n-2} \\ \text{Vâr}(\hat{\beta}_0) &= \frac{x_n x_1 s^2}{x_n - x_1} \\ \text{Vâr}(\hat{\beta}_1) &= \frac{s^2}{x_n - x_1} \\ \text{Côv}(\hat{\beta}_0, \hat{\beta}_1) &= \frac{-x_1 s^2}{x_n - x_1} \end{aligned} \quad (15.27)$$

While the slope estimate $\hat{\beta}_1 = (y_n - y_1)/(x_n - x_1)$ appears to make no use of the intermediate data $(y_2, x_2), \dots, (y_n, x_n)$, this is not really the

case because all values larger than y_i would not be what they are if a particular y_i had not been observed. The intermediate data are used, but they are used implicitly and their information is manifest only in the cumulative value y_n .

Consider the situation where $y = \beta x + \epsilon$, $E(y) = \beta x$, and the user has to choose between model A, the constant variance, uncorrelated errors model; and model B, the cumulative errors model. Mandel (1957) studied the effects of choosing the wrong error model upon the slope estimate and upon the estimated variance of the slope estimate. If the error is cumulative but model A is incorrectly chosen, then the user calculates the slope estimate and the estimated variance of the slope estimate with formulas derived from the wrong model. In this case the slope estimate is unbiased and even relatively efficient, but the estimate of the variance of the slope estimate greatly underestimates (on the average) the true variance of the slope estimate. This leads to confidence limits that are too small and tests of hypotheses with high probabilities of Type I error. In the second case the cumulative error model is assumed, but model A is correct. The slope, residual variance, and slope estimate variance are all calculated by formulas from the model B case. In this case the slope estimate is unbiased but very inefficient, and the variance of the slope estimate of the incorrectly chosen model is overestimated. This produces inflated confidence limits and hypothesis tests with little power.

It is clear that there are pitfalls associated with the misuse of each model and that it is important to choose the error model correctly. Some indication as to whether the errors are independent can often be obtained if y is plotted against x . The data will tend to vary systematically rather than randomly about the theoretical or fitted line if the cumulative model is applicable. Figure 15.5 from Mandel (1957) illustrates this point. For 50 equally spaced values of x from 1 to 50, and 50 values of ϵ_i taken from a table of random normal deviates, the values $y_i = 3x_i + \epsilon_i$ (independent data) and $y_i = 3x_i + \sum_{j=i}^i \epsilon_j$ (dependent data) are plotted. Note that while systematic behavior may arise from dependent errors in the observations, such dependence is only one possible cause for systematic behavior.

As indicated, the cumulative error model may be particularly useful as a method of vessel calibration. The rationale for its use is as follows. Consider a vessel that can be emptied completely and has the same cross-sectional area at all heights. The vessel is filled by independently weighing the nominally equal amounts added to the vessel. The corresponding manometer reading (here, the x value) is read exactly. If the errors in the weight additions are independent with equal variances (in the case of equal additions) or with variances proportional to the nominal weights added, then a cumulative error model of the first degree is appropriate. In practice, the process vessel may have to be divided into different regions

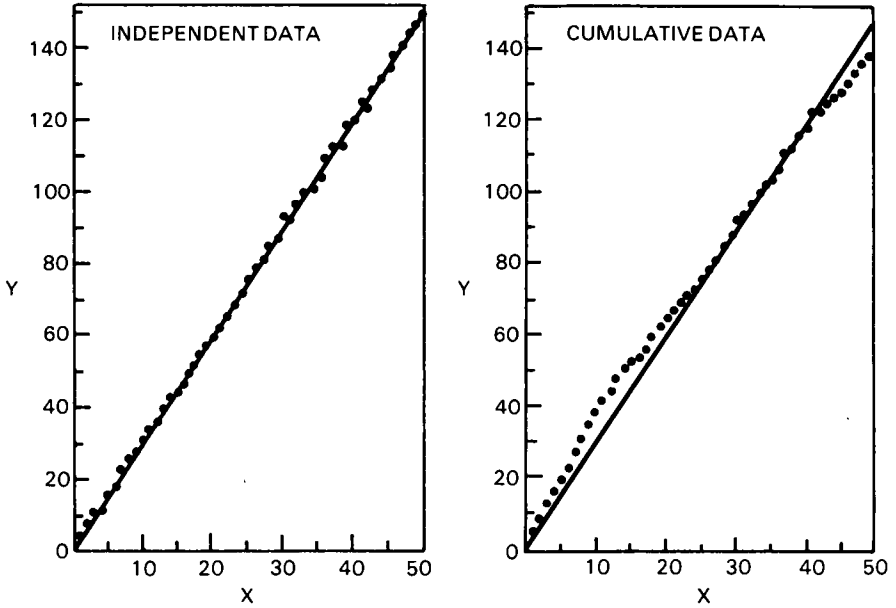


Figure 15.5 Comparison of independent vs. cumulative data: $n = 50$ (from Mandel, 1957, with permission).

because of heels originally present and because the cross-sectional areas are not uniform throughout the tank. The tank does not necessarily have to be a perfect cylinder where pipes and equipment do not intrude, but within a region these cross-sectional areas should be the same. Determining just where the cross-sectional areas change can be difficult if based only on the calibration data. The presence of heels can be accommodated by the use of a linear model containing a constant term.

One advantage that the cumulative model has when different regions of the vessel require different linear models is that the prediction line for the cumulative error model goes through the end points. Suppose that one region goes from (x_i, y_i) to (x_j, y_j) , $j > i$, and the next region goes from (x_j, y_j) to (x_k, y_k) , $k > j$. Then the prediction lines

$$y - y_i = \frac{y_j - y_i}{x_j - x_i} (x - x_i), \quad x_i < x < x_j$$

$$y - y_j = \frac{y_k - y_j}{x_k - x_j} (x - x_j), \quad x_j < x < x_k$$

have the point (x_j, y_j) in common.

Suppose that m calibration runs are made on a vessel that has no heel. From the cumulative model it follows that if $y_{n,1}, y_{n,2}, \dots, y_{n,m}$ (the last volume values for each run) are all at the same manometer reading x_n , then they should have the common variance $x_n \sigma^2$. Then the variance estimate

$$\hat{V}\text{ar}(y_n) = \left[\sum y_{n,i}^2 - \left(\sum y_{n,i} \right)^2 / m \right] / (m-1)$$

can be compared to

$$s^2 = x_n (s_1^2 + s_2^2 + \dots + s_m^2) / m$$

where s_i^2 is the residual variance from the i^{th} fitting. If $\hat{V}\text{ar}(y_n)$ is significantly larger than s^2 , the cumulative model does not hold. The more interesting situation is that in which $\hat{V}\text{ar}(y_n)$ is significantly smaller than s^2 . This result indicates that the fluctuations from calibration point to calibration point are due mostly to inherent variations in the process vessel and not in the weights of liquid added, and that the uncertainties associated with the cumulative model greatly overestimate the true uncertainty associated with each prediction point. Suppose that both manometer readings and measurements are made without error. Then in the absence of transients that affect the shape of the process vessel, the graphs of the calibration data (as irregular as they may appear to be) should be perfectly superimposable on each other.

The cumulative error model can be extended to higher order polynomials simply by differencing as before to make the errors uncorrelated and dividing through by $\sqrt{x_i - x_{i-1}}$ to make the variances equal. Then multiple regression fitting procedures and all the associated tests and results can be applied to the vectors $(z_i, t_{i,1}, t_{i,2}, \dots, t_{i,k})$, $i = 1, 2, \dots, n$.

15.1.5 Inverse Calibration

In the preceding sections, the calibration function has been estimated by considering the known values of the quantity to be estimated as the independent variable, and the observed instrumental or procedural responses as the dependent variable. However, as has been noted, it is the inverse of this calibration function that is used to determine the measurement result from an observed response on a process item. For linear calibration, this use of the calibration equation in the inverse form poses no particular problem. Further, the direct determination of the variance of the inverse estimate using the techniques illustrated in previous sections of this chapter is also relatively straightforward for quadratic functions, as

shown in Section 15.1.3.1 and by Jaech (1973). For cubic or higher-order polynomials, however, it is difficult or even impossible to obtain an explicit expression for the inverse function. The estimate itself and its approximate standard error may be relatively easy to determine by numerical or graphical means, but it is difficult to study other distributional and statistical properties in the absence of an explicit inverse equation.

An attractive approach to this problem is simply to invert the regression process, using the known values of the calibration standards as dependent variables and fitting them to an appropriate function of the measured responses. In an early article, Eisenhart (1939) argued that the direct, or "classical," approach should be preferred. More recently, arguments have been presented that the inverse approach should be preferred on its own merits, not simply because it is convenient to use (Krutchkoff, 1967). This has stimulated a great deal of discussion and numerous attempts to justify both "classical" and "inverse" methods on theoretical grounds. A recent review of the application of calibration procedures to tank calibration (Mullen, 1982) summarizes the arguments and gives extensive references. Although the classical approach as presented in this chapter seems to be favored, certainly no consensus has been reached.

Where a calibration function involving some form other than a simple first or second degree polynomial is required, approximate techniques can be applied to obtain the variances of the estimates. These techniques are described by Mullen (1982) and are illustrated in Section 15.1.3.1. The techniques are similar to those which have been used in nonlinear estimation problems and should be used when the number of parameters in the calibration function is greater than two or three.

15.2 MEASUREMENT CONTROL PROCEDURES

The use of known standards to obtain estimates of measurement error parameters and to monitor measurement processes for bias is one of the basic requirements of measurement control programs. The use of standards to monitor the measurement error parameters will be discussed in this section. To monitor measurement systems for bias, it is usually necessary to make measurements on "known" standards. The significant characteristic of such standards is that their values are assigned by a process that produces error effects which are negligible compared to the constant effects present in the measurement system being monitored. In some instances, it is only required that the systematic difference between the results from two measurement procedures (or two applications of the same procedure) be negligible—i.e., that the procedures are not biased relative to one another. With such a relaxed requirement, process materials may be used in place of standards.

There are a number of difficulties associated with the practice of estimating measurement error variances using data from the measurement of standards. This is particularly true of programs involving the destructive analysis of process materials to determine element or isotope concentrations. Standard materials have not always been available for all measurement applications, or their costs have been a limiting factor in their widespread usage. Also, known standards often are not completely representative of the types of materials actually being measured from the production process. There may be a number of factors, such as impurity content levels or stability of the standard, that can affect the magnitudes of the errors differently when measuring production samples and known standards. It is often difficult to adequately disguise a standard when it is submitted for analysis, and even the most conscientious analyst has problems being completely objective when it is known that the sample is not typical of the process materials. Further, it may be obvious to the analyst that certain steps in the measurement procedure—e.g., purification steps—are not required when measuring standard materials, thus eliminating potential sources of bias and random error in the final reported result. Consequently, error variances often tend to be underestimated when data are obtained from the measurement of standards.

Nevertheless, the use of known standards does play a very important role in the monitoring of measurement systems, as illustrated in Appendix 14A. Some measurement systems can be designed so the constant effects are small enough to be of no practical importance. For many measurement systems, however, they cannot be ignored and frequently they do not remain constant over time or changing conditions of measurement. In these cases, a measurement control program is needed that standardizes the conditions under which measurements are made and monitors the effect of changing conditions over time.

15.2.1 Monitoring Measurement Processes Against Standards

In-plant control programs are usually based on a planned program for the measurement of one or more standards. The replicate measurements can be used to determine the need for bias corrections, estimate the correction factor or factors to be applied, and estimate the measurement error variance components required to determine the variability in reported results.

15.2.1.1 An Example of Routine Monitoring

An example of data from enrichment measurements obtained over time on a standard will illustrate the methods for analyzing data which have been collected to monitor bias.

Example 15.7 Four NDA measurements of the percentage of ^{235}U in a sealed UO_2 standard having an assigned ^{235}U content $\mu_0 = 3.046\%$ with standard deviation $\sigma_0 = 0.0006$ are made each day. The data for 10 successive days are listed in the following table.

| Observation | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|-------------|--------|--------|--------|--------|--------|
| 1 | 3.025 | 3.033 | 3.049 | 3.006 | 3.010 |
| 2 | 3.022 | 3.044 | 3.041 | 2.978 | 3.050 |
| 3 | 3.017 | 3.047 | 3.075 | 3.022 | 3.039 |
| 4 | 3.024 | 3.049 | 3.078 | 3.028 | 3.057 |
| Total | 12.088 | 12.173 | 12.243 | 12.034 | 12.156 |
| Mean | 3.022 | 3.043 | 3.061 | 3.008 | 3.039 |

| Observation | Day 6 | Day 7 | Day 8 | Day 9 | Day 10 |
|-------------|--------|--------|--------|--------|--------|
| 1 | 2.998 | 3.048 | 2.984 | 3.042 | 3.037 |
| 2 | 3.002 | 3.072 | 2.972 | 3.017 | 3.051 |
| 3 | 3.024 | 3.108 | 3.002 | 3.053 | 3.034 |
| 4 | 3.013 | 3.057 | 2.954 | 3.044 | 3.039 |
| Total | 12.037 | 12.285 | 11.912 | 12.156 | 12.161 |
| Mean | 3.009 | 3.071 | 2.978 | 3.039 | 3.040 |

| | | | |
|-------------|---------|--------------|----------|
| Grand total | 121.245 | Overall mean | 3.031125 |
|-------------|---------|--------------|----------|

An appropriate model for these control data is analogous to the error model developed for the bulk measurement calibration in Appendix 14A. Three levels of error effects are assumed: deviations from a daily mean value, deviations of the daily mean from an overall mean for the data set, and a deviation of the overall mean from the standard value. The model for a one-way analysis of variance appropriate for these data is given by Equation 5.6 of Section 5.1.1, and the appropriate analysis for determining variance components is given in Section 5.2.1. Because the daily sample sizes are equal, $n_0 = n = 4$, and there are $a = 10$ days in the analysis, so that $N = an = 10(4) = 40$. From Table 5.3, Section 5.2.3, and the calculational methods given in Section 5.1.1, the analysis of variance given in Table 15.13 can be computed. This analysis of variance incorporates the

TABLE 15.13
Analysis of Variance for Routine Monitoring Data

| Source | df | SS | MS | E(MS) | F |
|-------------------------|----|----------|-----------|---|-------|
| Deviation from standard | 1 | 0.008851 | 0.008851 | $\sigma^2 + 4\sigma_\alpha^2 + 40(\mu - \mu_0)^2$ | 2.955 |
| Between days | 9 | 0.026952 | 0.002995 | $\sigma^2 + 4\sigma_\alpha^2$ | 10.44 |
| Within days | 30 | 0.008615 | 0.0002872 | σ^2 | |

test for an overall bias in this data set. The sum of squares for the single deviation $-0.014875 = 3.031125 - 3.046$ of the overall mean from the standard is given by $40(-0.014875^2) = 0.008851$. The remaining sums of squares are obtained from Equations 5.19 and 5.20 using the sums

$$S_1 = \sum_1^{40} x_{ij}^2 = 367.544317$$

$$S_2 = \sum \frac{T_i^2}{4} = \frac{1470.142809}{4} = 367.53570225$$

$$S_3 = \frac{T^2}{40} = \frac{121.245^2}{40} = 367.508750626$$

where T_i and T are the daily totals and grand total, respectively. Using the mean squares between (MS_B) and within (MS_W) as estimates of their respective expected values, which are linear combinations of the variance components of interest, the estimates

$$\hat{\sigma}^2 = 0.0002872, \hat{\sigma} = 0.0169$$

and

$$\hat{\sigma}_\alpha^2 = \frac{0.002995 - 0.0002872}{4} = 0.0006769, \hat{\sigma}_\alpha = 0.0260$$

are obtained for the variability within and between days. The calculated F for testing the hypothesis that there is no difference in the day-to-day deviations from the overall mean is 10.44, which is significant at the 0.01 level for 9 and 30 degrees of freedom. Therefore, there is good reason to conclude that the variability in the measurement method from day to day is greater than the variability within a given day. This is also quite apparent from Figure 15.6 where the data are plotted sequentially.

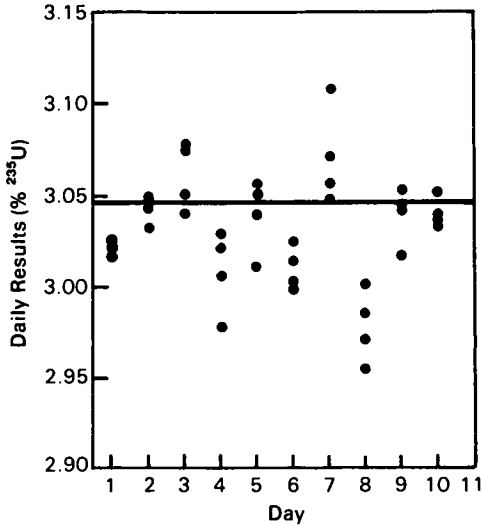


Figure 15.6 Daily results of enrichment measurements on a standard.

The F-test for the deviation from the standard value is a test of the null hypothesis that there is no long-term constant error effect in this data set. There are two possible tests for this hypothesis, depending on whether the interest is only in the 10 days represented by the data in the example, or whether the 10 days are to be considered as a sample from a larger population of days and the interest is in making inferences about this larger population (Section 5.2.3). The calculated F in the analysis of variance table is for the second situation and is obtained by forming the ratio MS_M/MS_B . The resulting value of 2.955 is not significant even at the 0.10 level, and there is insufficient evidence to conclude that a long-term effect exists. For the alternate situation where the inferences are to apply to only the 10 days represented by this particular set of data, the effects α_i are fixed and the second term in the expected mean square for the deviation from the standard value drops out, so that the F-ratio would be calculated as MS_M/MS_W . This second situation is not the usual one for most safeguards applications.

The following is an intuitive explanation for the difference in the expected mean squares. When the days represented in the data are considered to be a sample from a larger population, there is uncertainty in $\bar{x}_..$ due to the finite sample of days. Even if there were no measurement variability within days—i.e., no deviations from the daily means—another sample of days from the larger population would be expected to give rise to a different estimate of the mean. However, when the sampled days represent all of the days in the population of interest, the estimate of the mean is fixed except for random measurement error.

It is important from a practical point of view to compare the uncertainty in measurement results when a bias correction has been made with the alternative of reporting the result without a bias correction being made. In addition to the technical problems there are almost always practical considerations, including the bookkeeping difficulties associated with making corrections to results which have already been posted and the need to allocate both money and effort over the entire spectrum of measurement methods used for safeguards purposes. For some measurement methods, there will be concern for only a single bias to be estimated and the need for correction determined. Such a bias is represented by a single error effect that is constant over the entire time period (or data set) of interest. As indicated by the analysis of this example, however, there are usually additional error effects which change with day-to-day operations, calibration periods, instruments, or other factors.

For the model involving a single intermediate source of variability used to represent the above data from routine monitoring, the mean square error (MSE) of a single future measurement y , uncorrected for bias, is given by

$$\text{MSE}(y) = [(\mu - \mu_0)^2 + \sigma_0^2] + [\sigma^2 + \sigma_\alpha^2] \quad (15.28)$$

As pointed out in Appendix 14A, the problem in using this equation is that even if we assume the standard value μ_0 to be unbiased, the expected value of $(\mu - \mu_0)^2$ for a given data set must either be known from historical experience or estimated from a mean square with only one degree of freedom (see Table 15.13). If a correction is made based on the difference between the overall mean of the data set and the assigned value of the standard, the bias term drops out of the mean square error if the assigned value of the standard is unbiased. The variance of a corrected measurement is given by

$$\text{Var}(y) = \left[\frac{\sigma_\alpha^2}{a} + \frac{\sigma^2}{N} + \sigma_0^2 \right] + [\sigma^2 + \sigma_\alpha^2] \quad (15.29)$$

This variance can be estimated from the known variance of the standard and the estimates $\hat{\sigma}^2$ and $\hat{\sigma}_\alpha^2$ obtained from the data analysis. When the difference between each daily mean and the standard value is used to correct each day's results, the variance of a single future measurement result is

$$\text{Var}(y) = \frac{\sigma^2}{n} + \sigma_0^2 + \sigma^2 \quad (15.30)$$

The contribution of the variance component σ_α^2 has dropped out of this variance, but the contribution to the variability in the correction factor from the within-days variability, σ^2/n , is a factor of a larger than the contribution σ^2/N from this source in Equation 15.29.

For the data of this example, a single correction to be added to future results would be the difference $3.046 - 3.031125 = 0.0149$ between the assigned value of the standard and the overall mean. Substituting the estimates of σ^2 and σ_α^2 obtained from the analysis of variance in Table 15.13 and the assigned standard error σ_0 of the standard into Equation 15.29, an estimate of the variance of a single future measurement corrected in this fashion is

$$\begin{aligned} \text{V}\hat{\text{a}}\text{r}(y) &= \left[\frac{0.0006769}{10} + \frac{0.0002782}{40} + (0.0006)^2 \right] \\ &\quad + [0.0006769 + 0.0002782] \\ &= 0.0000750 + 0.0009551 \\ &= 0.0010301 \end{aligned}$$

and

$$\hat{\text{S}}\hat{\text{E}}(y) = 0.0321$$

The contribution of the variance of the assigned value of the standard is negligible, and the contribution of the variance of the correction factor is relatively small. If daily corrections are made based on the difference between the assigned value of the standard and each daily mean (e.g., adding an amount $3.046 - 3.022 = 0.024$ to all production measurements made on day 1) the estimated variance is computed from Equation 15.30 as

$$\begin{aligned} \text{V}\hat{\text{a}}\text{r}(y) &= \left[\frac{0.002872}{4} + (0.0006)^2 \right] + 0.0002872 \\ &= 0.0000722 + 0.0002872 \\ &= 0.0003594 \end{aligned}$$

and

$$\hat{\text{S}}\hat{\text{E}}(y) = 0.0190$$

The largest decrease in this estimate is due to the elimination of the between-day variability from the measurement itself. The contribution of the variance of the assigned value is still negligible. The results of this analysis indicate that to keep both the estimate of the inventory difference unbiased and the uncertainty in the estimate small, it is desirable to routinely correct future measurement results using an estimate of the daily bias. Making a single bias correction for the 10-day period does not significantly improve the uncertainty.

The pattern of the data in Figure 15.6 is not at all unusual for measurements on a standard made over time. Data of this type often show apparent structure such as time trends, cyclical movements, abrupt shifts, or other nonrandom behavior. When data are available for a sufficiently long period of time, it may also be possible to demonstrate that there is a significant average bias for the entire period. If the data used in this example are unusual, it is because the daily averages do not appear to show any nonrandom patterns, although 10 days is a relatively short time for such patterns to occur.

Figure 15.7 shows the daily averages plotted on a control chart, which can be used as one tool for routinely monitoring measurement systems for

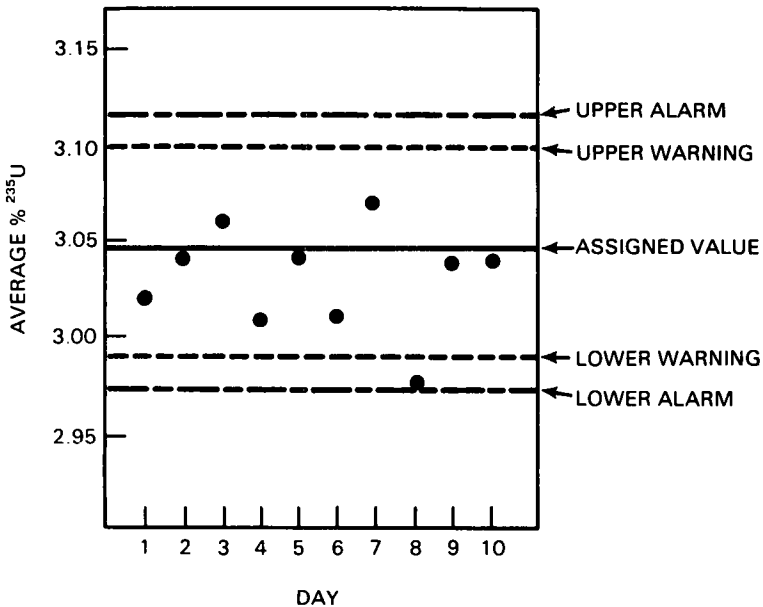


Figure 15.7 Control chart for daily averages of measurements on a standard.

bias. Construction of the control chart, using the procedures given in Section 4.7.1.1 and based on the normal distribution, proceeded as follows:

1. The central (solid) line was drawn at the assigned value, $\mu_0 = 3.046$, of the standard.
2. The estimated standard deviation of an observed daily average \bar{y}_i based on four measurements per day was calculated using

$$\hat{SE}(\bar{y}_i) = \left(\hat{\sigma}_\alpha^2 + \frac{\hat{\sigma}^2}{4} + \sigma_0^2 \right)^{1/2} = 0.02737$$

The estimates $\hat{\sigma}$ and $\hat{\sigma}_\alpha$ obtained in this example were used in this calculation. These estimates are sometimes obtained from larger sets of available data if the measurement method has been shown to be stable over the longer period.

3. The warning limits were calculated using

$$\begin{aligned} \text{Upper warning} &= \mu_0 + 1.96 \hat{SE}(\bar{y}_i) = 3.100 \\ \text{Lower warning} &= \mu_0 - 1.96 SE(\bar{y}_i) = 2.992 \end{aligned}$$

where 1.96 is obtained from Table A3 so that the probability is 0.95 that an observation will fall between the warning limits when the process is in control.

4. The alarm limits were calculated using

$$\begin{aligned} \text{Upper alarm} &= \mu_0 + 2.58 \hat{SE}(\bar{y}_i) = 3.117 \\ \text{Lower alarm} &= \mu_0 - 2.58 SE(\bar{y}_i) = 2.975 \end{aligned}$$

where 2.58 is obtained from Table A3 so that the probability is 0.99 that an observation will fall between the alarm limits when the process is in control.

The use of these warning and alarm limits, instead of the three sigma limits more common in quality control procedures (Section 4.7.1.1), is based on standardized NRC procedures. When the approximate degrees of freedom in the variance estimates are known to be small, it may be desirable to compute these limits using the appropriate values from the Student's *t*-distribution (Table A5) rather than the normal distribution (Section 5.3.1). The measurement control data for the 10 days in the example data set show no evidence of a serious lack of control or an appreciable constant bias over the time period. Because one average is below the lower warning

limit and 8 of the 10 averages are below the assigned value of the standard, there is some evidence that a small bias may be present, but more data are needed to confirm this.

Data gathered to estimate or monitor bias also frequently contain information about the random error variance of the measurement method. As discussed in Section 15.1, however, there may be reasons why the random variance estimates calculated from data obtained on standards underestimate the actual random variability when process materials are measured. Known standards often are not completely representative of the types of process samples actually being measured. The degree to which the standard fails to represent the process items or samples may vary widely. When the measurement is a chemical analysis, the presence or absence of certain impurities in the sample being analyzed can have an effect on the error in the final result not found in the measurements on a standard. When possible, it is desirable to obtain random error variance estimates from independent, repeated measurements on actual process samples. However, when a set of standards measurement data collected over time is being analyzed for other purposes, such as testing for changes in the measurement bias, the added effort of monitoring the random error variance for measurements on the standard may be worthwhile.

Suppose that the analysis of a large amount of historical data for the NDA instrument used to measure enrichment has resulted in a standard deviation estimate of $\hat{\sigma} = 0.015$ when measuring UO_2 standards with assigned percent ^{235}U values close to 3%. A relatively easy way to routinely monitor the random variability for this measurement procedure is through the use of a control chart for the range or the standard deviation. Procedures for constructing both types of control charts are given in Sections 4.7.1.2 and 4.7.1.3, but only the range chart will be illustrated here. Using the values of D_1 and D_2 from Table 4.3, the upper and lower three-sigma control limits for the range when $n = 4$ are computed as

$$\text{UCL} = D_2\hat{\sigma} = 4.698(0.015) = 0.0705$$

$$\text{LCL} = D_1\hat{\sigma} = 0(0.015) = 0$$

Alternatively, the upper 5% and 1% points for the distribution of the range when $n = 4$ are 3.63σ and 4.40σ , respectively (Bennett and Franklin, 1954, Table 10.7), from which the corresponding warning and alarm limits are calculated as $(3.63)(0.015) = 0.0544$ and $(4.40)(0.015) = 0.0660$. The observed ranges are as follows:

| Day | Range |
|-----|-------|
| 1 | 0.008 |
| 2 | 0.016 |
| 3 | 0.037 |
| 4 | 0.050 |
| 5 | 0.047 |
| 6 | 0.026 |
| 7 | 0.060 |
| 8 | 0.048 |
| 9 | 0.036 |
| 10 | 0.017 |

These observed ranges are plotted on the control chart shown in Figure 15.8. All are within the control limit and the alarm limit. One point is outside the warning limit, but this is not unusual for 10 points. While this procedure does not provide conclusive evidence that the measurement method was performing in a controlled state with respect to the random variability when measuring process samples, it is one supporting type of information because the performance while measuring the standard appears acceptable.

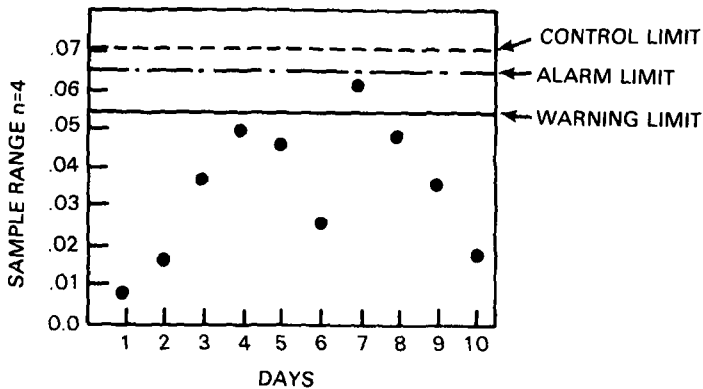


Figure 15.8 Control chart for range of four daily measurements of a standard.

15.2.1.2 Continuous Monitoring Against a Standard

When intermediate-level variability is present as in the preceding example, it is desirable to monitor and correct measurement results continuously if this process reduces the estimated variability in the process

measurements. The basic models and estimates associated with such correction procedures are given in this section. The assumed models are

$$\begin{aligned}x_i &= \mu_0 + \beta + \epsilon_i \\y_j &= T_j + \beta + \eta_j\end{aligned}$$

where

- x_i = i^{th} measurement on the standard, $i = 1, \dots, n$
- y_j = measurement on the j^{th} production item
- T_j = true value of the characteristic for the j^{th} item
- β = common fixed effect for both standard and item measurements
- ϵ_i = individual error in standard measurement with variance σ_ϵ^2
- η_i = individual error in item measurement with variance σ_η^2
- μ_0 = assigned value for the standard.

The estimate of the bias is

$$\hat{\beta} = \bar{x} - \mu_0 \quad (15.31)$$

Let the bias-corrected result for the j^{th} production item be

$$y'_j = y_j - \hat{\beta} \quad (15.32)$$

Then the variance component associated with using the estimated bias is given by

$$\text{Var}(\hat{\beta}) = \frac{\sigma_\epsilon^2}{n} + \sigma_0^2 \quad (15.33)$$

where σ_0^2 is the variance in the assignment of the standard value, and the total variability in the bias-corrected result for the j^{th} item is

$$\text{Var}(y'_j) = \sigma_\eta^2 + \frac{\sigma_\epsilon^2}{n} + \sigma_0^2 \quad (15.34)$$

The covariance between y'_j and y'_k when the same bias correction is used is

$$\text{Cov}(y'_j, y'_k) = \frac{\sigma_\epsilon^2}{n} + \sigma_0^2 \quad (15.35)$$

Measurements using different bias corrections are independent.

When the bias correction is not made on the measurement results for production items, the mean square error can be used as a measure of uncertainty. The mean square error for y_j is

$$\text{MSE}(y_j) = \sigma_\eta^2 + \beta^2 \quad (15.36)$$

In most practical situations, the true values of neither the bias nor σ_η^2 will be known. The estimate $\hat{\beta}$ can be used for β and the estimate s^2 for σ_η^2 , where s^2 is the sample variance for the n standard measurements. However, not only is $\hat{\beta}$ highly variable when estimated in this manner, but also the expected value of $\hat{\beta}^2$ is not equal to β^2 . Rather

$$E(\hat{\beta}^2) = \beta^2 + \frac{\sigma_\epsilon^2}{n} + \sigma_0^2 \quad (15.37)$$

There appear to be three reasonable options for estimating β^2 :

1. Use $\hat{\beta}^2$ (a biased estimator).
2. Use $\hat{\beta}^2 - \frac{s^2}{n} - \sigma_0^2$ (an unbiased estimator).
3. Use $\hat{\beta}^2 - \frac{s^2}{n} - \sigma_0^2$ whenever positive; use zero otherwise.

The third option is a compromise which, although the estimator is biased, does eliminate the problem of negative estimates and has a smaller variance than the other two alternatives. From the standard measurement data alone there is no direct estimate of σ_η^2 , but it is frequently a reasonable assumption that the random variability in measurements is the same for production items and standards, so that s^2 can be used. When feasible, this assumption should be checked by some replication of production measurements.

An alternative to the extremes of always correcting for bias or never correcting for bias is to correct only when a significant bias exists at some significance level α . The argument for this procedure is based on the fact that if β^2 is small compared to $\sigma^2 = \sigma_\epsilon^2 = \sigma_\eta^2$, then the advantage of eliminating the bias term β^2 from the mean square error may be completely negated by the introduction of a large contribution σ^2/n into the uncertainty of the bias-corrected results. In addition, the problems associated with estimating β^2 are greatest when it is small. Stewart (1975) has studied the ratio of the mean square error under the two extreme procedures (disadvantage factor) as a function of the ratio β/σ , and gives recommendations for optimal procedures.

15.2.1.3 Procedures Involving Several Standards

There are several situations in which more than one standard will be measured for purposes of estimation and monitoring of a measurement method. Two of the more common applications are described in this section. In the first type of application, a given scale is used to measure both the gross and the tare weights of a container to determine the net weight of material in the container. It is necessary to evaluate the bias for the scale in both the gross and tare weight ranges in order to estimate the bias for a reported net weight. In the second application, more than one standard is measured covering either the calibration range or a more limited range of actual usage for the measurement method.

In the general case where there is no pattern to the biases over the range of interest, the problem must be handled as a series of single standard studies (as presented in previous sections). When a pattern for the bias over the range of usage—such as a constant bias or constant relative bias—cannot be established through analysis of available data, bias estimates need to be obtained in each of the subranges in which the measurement method will be used. It is assumed in this section that any existing bias is constant, on either an absolute or relative basis, over the range covered by the standards. For both applications, different standards may be measured a different number of times during the period of interest and the standards may have different variances associated with their assigned values for the characteristic of interest.

The following notation will be used for the first application:

- μ_g = true value of the gross weight standard
- \bar{z}_g = assigned value of the gross weight standard
- σ_g^2 = variance of the assigned gross weight standard value
- μ_t = true value of the tare weight standard
- \bar{z}_t = assigned value of the tare weight standard
- σ_t^2 = variance of the assigned tare weight standard value
- \bar{x}_g = average of n_g measurements on gross weight standard
- \bar{x}_t = average of n_t measurements on tare weight standard
- s_g^2 = sample variance of the measurements on the gross weight standard
- s_t^2 = sample variance of the measurements on the tare weight standard
- y_j = measured net weight for production item j

The estimated bias for a net weight determination is given by

$$\hat{\beta} = (\bar{x}_g - \bar{z}_g) - (\bar{x}_t - \bar{z}_t) \quad (15.38)$$

An estimate of the variance of this estimated bias is

$$\text{Vâr}(\hat{\beta}) = \hat{\sigma}_g^2 + \hat{\sigma}_t^2 + \frac{s_g^2}{n_g} + \frac{s_t^2}{n_t} \quad (15.39)$$

where the first two variances in this expression are estimates that accompany the standards or have been established on other grounds. If the two sample variances are valid estimates of the measurement variances when actual containers of material are weighed, the total variance of a bias corrected net weight for a container of material is estimated using

$$\text{Vâr}(y'_j) = s_g^2 + s_t^2 + \hat{\sigma}_g^2 + \hat{\sigma}_t^2 + \frac{s_g^2}{n_g} + \frac{s_t^2}{n_t} \quad (15.40)$$

where the bias corrected result is given by

$$y'_j = y_j - \hat{\beta} \quad (15.41)$$

If the bias correction is not made, the mean square error of the reported result, y_j , is slightly more complicated than was the case for a single standard (Equation 15.36). An estimate of the mean square error is

$$\text{M\hat{S}E}(y_j) = (s_g^2 + s_t^2) + \left(\hat{\beta}^2 - \hat{\sigma}_g^2 - \hat{\sigma}_t^2 - \frac{s_g^2}{n_g} - \frac{s_t^2}{n_t} \right) \quad (15.42)$$

The same concerns with the estimation of β^2 are present as those discussed in the example in the previous section.

Example 15.8 The standard values assigned to the gross weight and tare weight standards referred to in Example 15.2 were:

$$\bar{x}_g = 8878.0 \text{ g}$$

$$\bar{x}_t = 1591.7 \text{ g}$$

Assume $\hat{\sigma}_g = \hat{\sigma}_t = 0.97 \text{ g}$ and that check weighings of these standards yield the following information:

$$n_g = 30$$

$$\bar{x}_g = 8881.3 \text{ g}$$

$$s_g = 6.2 \text{ g}$$

$$n_t = 20$$

$$\bar{x}_t = 1589.9 \text{ g}$$

$$s_t = 4.9 \text{ g}$$

From these results the estimated bias (Equation 15.38) for a net weight is computed as

$$\begin{aligned}\hat{\beta} &= (8881.3 - 8878.0) - (1589.9 - 1591.7) \\ &= 3.3 - (-1.8) = 5.1\end{aligned}$$

If the bias estimate is applied as a correction to a net weight measurement for a process item, an estimate of the variance of the corrected value is given by Equation 15.40 as

$$6.2^2 + 4.9^2 + (0.97)^2 + (0.97)^2 + (6.2)^2/30 + (4.9)^2/20 = 66.8136$$

If the bias estimate is not applied, an estimate of the mean square error is given by Equation 15.42 as

$$(6.2^2 + 4.9^2) + [5.1^2 - 0.97^2 - 0.97^2 - (6.2)^2/30 - (4.9)^2/20] = 84.0964$$

Because the estimated variance of the corrected value is smaller than the mean square error estimate for the uncorrected value, it probably would be advisable either to correct net weight measurements on process items or to perform maintenance and recalibration of the scale to reduce the bias.

For the second application, the following notation will be used:

- m = number of standards used to estimate bias
- n_k = number of measurements on standard k
- μ_k = true value for standard k
- \bar{z}_k = assigned value for standard k
- $\sigma_{0,k}^2$ = variance of assigned value for standard k
- \bar{x}_k = average of the measurements on standard k
- σ_k^2 = variance of the measurements on standard k
- s_k^2 = sample variance of the measurements on standard k

Because both the number of measurements, the variance of the assigned value of the standard, and the variance of the measurements may be different from standard to standard, a weighted average should be used to estimate the bias, so that

$$\hat{\beta} = \frac{\sum_{k=1}^m w_k(\bar{x}_k - \bar{z}_k)}{\sum_{k=1}^m w_k} \quad (15.43)$$

where

$$w_k = \frac{1}{(\sigma_{0,k}^2 + \sigma_k^2/n_k)} \quad (15.44)$$

The correct values of w_k are functions of the two unknown variances. If the standards being used for the bias estimation originated with a national or international standards organization, the certification statements that accompany the assigned values for the standards usually give estimates $\hat{\sigma}_{0,k}^2$ of $\sigma_{0,k}^2$. For other standards and for the measurement error variances, sample estimates are needed to ensure that the weights used are approximately correct. If it can be demonstrated that the measurement variances are approximately the same for all standards—i.e., tests show that the hypothesis $\sigma_k^2 = \sigma^2$ for all k is not rejected—then a pooled estimate, s^2 , can be calculated using the measurement data obtained on all of the m standards. The pooled estimate of the variance is given by

$$s^2 = \sum_{k=1}^m (n_k - 1)s_k^2 / (N - m) \quad (15.45)$$

where $N = \sum_k n_k$ is the total number of measurements on all standards.

The variance of the estimate $\hat{\beta}$ is $\left(\sum_{k=1}^m w_k \right)^{-1}$. If s^2 is a valid estimate of the measurement variance of measurements on production items as well as standards, then an estimate of the variance of a corrected measurement on a production item is given by

$$V\hat{a}r(y_j) = s^2 + \left(\sum_{k=1}^m \hat{w}_k \right)^{-1} \quad (15.46)$$

where \hat{w}_k is computed from Equation 15.44 using $\sigma_{0,k}^2 = \hat{\sigma}_{0,k}^2$ and $\sigma_k^2 = s^2$. The mean square error of an uncorrected measurement on a production item is estimated as

$$M\hat{S}E(y_j) = s^2 + \hat{\beta}^2 - \left(\sum_{k=1}^m \hat{w}_k \right)^{-1} \quad (15.47)$$

Example 15.9 Three standards are used in a laboratory to monitor a mass spectrometric method for bias in the measurement of percent ^{235}U . The assigned values for the three standards in percent ^{235}U are 2.013, 3.009, and 4.949. The relative standard deviations for the assigned values are each equal to 0.05%, i.e., $\delta_{0,1} = \delta_{0,2} = \delta_{0,3} = 0.0005$. During the time period of concern, the following measurements on these standards were obtained:

| Standard 1 | Standard 2 | Standard 3 |
|------------|------------|------------|
| 2.013 | 3.008 | 4.953 |
| 2.017 | 3.013 | 4.957 |
| 2.015 | 3.010 | 4.949 |
| 2.013 | 3.017 | 4.946 |
| 2.011 | 3.009 | 4.951 |
| 2.015 | 3.008 | 4.958 |
| | 3.014 | 4.945 |
| | 3.010 | |
| | 3.011 | |
| | 3.006 | |

Past experience in the laboratory has demonstrated that for this mass spectrometric method, both the random errors and bias are constant on a relative basis. From the discussion in Section 14.2.1, it is appropriate to analyze these data using the logarithms of the measurements. The assigned values for the three standards in this transformed scale, using natural logarithms, are as follows:

$$\bar{z}_1 = \ln 2.013 = 0.699626$$

$$\text{Vâr}(\bar{z}_1) \cong \delta_{0,1}^2 = (0.0005)^2$$

$$\bar{z}_2 = 1.101608$$

$$\text{Vâr}(\bar{z}_2) \cong \delta_{0,2}^2 = (0.0005)^2$$

$$\bar{z}_3 = 1.599186$$

$$\text{Vâr}(\bar{z}_3) \cong \delta_{0,3}^2 = (0.0005)^2$$

since the absolute variance of the transformed values is approximately the relative variance of the original values. The sample averages and variances for the natural logarithms of the observed results are as follows:

$$\bar{x}_1 = 0.700122$$

$$s_1^2 = 1.084762 \times 10^{-6}$$

$$\bar{x}_2 = 1.102139$$

$$s_2^2 = 1.181034 \times 10^{-6}$$

$$\bar{x}_3 = 1.599647$$

$$s_3^2 = 1.042966 \times 10^{-6}$$

Because $m = 3$, $n_1 = 6$, $n_2 = 10$, and $n_3 = 7$, and, from Equation 15.45,

$$\begin{aligned} s^2 &= [5(1.084762) + 9(1.181034) \\ &\quad + 6(1.042966)] \times 10^{-6} / (23 - 3) \\ &= 1.115546 \times 10^{-6} \end{aligned}$$

the three weights can be estimated from Equation 15.44 as

$$\hat{w}_1 = \frac{1}{(0.0005)^2 + [(1.115546 \times 10^{-6})/6]} = 2.2940 \times 10^6$$

$$\hat{w}_2 = \frac{1}{(0.0005)^2 + [(1.115546 \times 10^{-6})/10]} = 2.7658 \times 10^6$$

$$\hat{w}_3 = \frac{1}{(0.0005)^2 + [(1.115546 \times 10^{-6})/7]} = 2.4428 \times 10^6$$

and

$$\sum_{k=1}^3 \hat{w}_k = 7.5026 \times 10^6$$

The estimated bias is then

$$\begin{aligned} \hat{\beta} &= (1/7.5026) [2.2940(0.700122 - 0.699626) \\ &\quad + 2.7658(1.102139 - 1.101608) \\ &\quad + 2.4428(1.599647 - 1.599180)] \\ &= 0.0037326/7.5026 = 0.0004975 \end{aligned}$$

Subtracting this estimated bias correction from the logarithm of the measured percent ^{235}U is equivalent to multiplying each measured result by 0.9995. If this correction is made for measurements of percent ^{235}U in production samples, and it is assumed that s^2 is an estimate of the relative measurement variance for routine analyses, the relative variance of corrected results is estimated using Equation 15.46 as

$$\begin{aligned}
 s^2 + \left(\sum_{k=1}^3 \hat{w}_k \right)^{-1} &= 1.1155 \times 10^{-6} + (7.5026 \times 10^6)^{-1} \\
 &= 1.1155 \times 10^{-6} + 0.1333 \times 10^{-6} \\
 &= 1.2488 \times 10^{-6}
 \end{aligned}$$

The estimate of mean square error for uncorrected results is, from Equation 15.47,

$$1.1155 \times 10^{-6} + (0.0004975^2 - 0.1333 \times 10^{-6}) = 1.2297 \times 10^{-6}$$

There is little justification for the routine application of a bias correction because the variances are essentially the same.

15.2.2 Interlaboratory Control Programs

In quite a few nuclear safeguards applications, more than one processing facility or analytical laboratory is involved in some phase in the handling of and the accounting for the nuclear materials. Most notably, this multifacility (laboratory) involvement occurs when shipper-receiver differences are being examined for one or more shipments of material between one facility and another. The need exists, therefore, for an extension to more than one laboratory of the procedures for estimating and monitoring bias for a measurement system which is operating under a fixed set of controlled conditions within one laboratory. In cases where multiple samples drawn from a single standard material (or solution) are distributed in interlaboratory exchange or "round robin" studies, the primary purpose for the exchange is often one of obtaining estimates of the measurement bias existing in the participating laboratories and to determine whether these estimates of bias deviate from zero or are different from one laboratory to another.

The data collection plan given to each laboratory participating in such round robin studies can vary from simple to complicated depending on the need to identify and assess the relative importance of the factors or steps in the measurement procedure that contribute to the total uncertainty of a measurement result. In simple cases, the purpose of the study may be restricted to estimation and evaluation of the biases present. A breakdown of the contributions to the bias and random error terms due to the various factors or steps in the measurement is not needed, and each laboratory might simply be instructed to obtain n replicate measurements of the sample received. The instructions may specify the measurement method to use,

the operating conditions under which the method is to be applied, and the format to be used in reporting the measurement results. A simple linear model for such data might be:

$$x_{ik} = \mu_0 + \beta_i + \epsilon_{ik}$$

where

- x_{ik} = the k^{th} measurement at the i^{th} laboratory
- $i = 1, 2, \dots, m$ laboratories
- $k = 1, 2, \dots, n$ replicate measurements
- μ_0 = the assigned reference value for the standard
- β_i = the bias effect for the i^{th} laboratory
- ϵ_{ik} = the random error for the k^{th} measurement at the i^{th} laboratory.

The data from a simple round robin study of this type can be analyzed using the appropriate one-way analysis of variance techniques described and illustrated in Section 5.1 or 5.2. Even when one or more of the participating laboratories does not obtain or report the requested number of replicate measurements, the data they do provide can be analyzed using a one-way analysis of variance for unbalanced data sets. The purpose of the analysis is to determine whether there is evidence of an overall bias in the measurement method which is common to all of the laboratories and whether there is evidence of differences in the biases of the laboratories.

15.2.2.1 Shipper-Receiver Bias

It is a common situation in round robin studies in safeguards that the particular laboratories participating in the study are the only laboratories of interest and constitute the entire population for which inferences are to be made. To illustrate, the three participating laboratories in a round robin study might be a shipper's laboratory, a receiver's laboratory, and a referee's (or inspector's) laboratory. The major objective of the study in this case would be to examine the possibility of a bias between the shipper's and receiver's laboratories. For such studies, the one-way fixed effects analysis of variance of Section 5.1 provides the appropriate methodology.

Example 15.10 Samples of NBS standard reference material 949d were sent to each of three participating laboratories: a shipper's laboratory, a receiver's laboratory, and a referee's laboratory. Each laboratory was instructed to analyze its sample 10 times for plutonium concentration

and to report the individual results. The receiver's laboratory reported only nine results. Table 15.14 summarizes the reported percent recoveries.

TABLE 15.14
Reported Percent Recovery

| Shipper
x_{1k} | Receiver
x_{2k} | Referee
x_{3k} |
|-----------------------|------------------------------|----------------------|
| 100.35 | 99.94 | 99.65 |
| 100.35 | 100.05 | 99.92 |
| 100.18 | 100.06 | 100.13 |
| 100.28 | 100.33 | 99.77 |
| 100.30 | 99.77 | 99.98 |
| 100.27 | 99.91 | 100.21 |
| 100.19 | 100.08 | 99.99 |
| 100.23 | 99.86 | 99.77 |
| 100.12 | 100.00 | 100.12 |
| 100.29 | | 99.88 |
| <hr/> | | |
| Totals = 1002.56 | 900.00 | 999.42 |
| $\bar{x}_1 = 100.256$ | $\bar{x}_2 = 100.000$ | $\bar{x}_3 = 99.942$ |
| $s_1^2 = 0.005649$ | $s_2^2 = 0.0257006$ | $s_3^2 = 0.032373$ |
| <hr/> | | |
| Grand total = 2901.98 | $\bar{x}_{..} = 100.0682759$ | |

Using the procedures of Sections 5.1.2 and 5.1.1, the calculations for the analysis of variance on these data, with $a = 3$, $n_1 = n_3 = 10$, $n_2 = 9$, and $N = 29$, are as follows:

$$SS_1 = \sum_k x_{ik}^2 = 100.35^2 + 100.35^2 + \dots + 99.88^2 = 290,397.2368$$

$$SS_2 = \sum \frac{T_i^2}{n_i} + = \frac{1002.56^2}{10} + \frac{900.00^2}{9} + \frac{999.42^2}{10} = 290,396.6890$$

$$SS_3 = \frac{T^2}{29} = \frac{2901.98^2}{29} = 290,396.135186$$

The resulting analysis of variance obtained from Equations 5.28 through 5.35 is

| Source | df | SS | MS | E(MS) | F |
|--------------|----|--------|---------|--|-------|
| Between labs | 2 | 0.5538 | 0.2769 | $\sigma^2 + \sum_{i=1}^3 \frac{n_i(\mu_i - \bar{\mu})^2}{2}$ | 13.14 |
| Within labs | 26 | 0.5478 | 0.02107 | σ^2 | |

where $\bar{\mu} = \frac{1}{N} \sum_{i=1}^3 n_i \mu_i$ and $\mu_i = \mu_0 + \beta_i$. If the test of significance is made at the 0.05 level, the critical value for the F-ratio with 2 and 26 degrees of freedom is 3.37. There is evidence of significant bias between the three laboratories based on this set of data.

In this example, the assumption of equal variances for the three sets of results could be questioned in view of the relatively small value of s_i^2 . A simple approximate test of the homogeneity of several variances is given in Bennett and Franklin (1954). The test statistic is

$$B = \frac{1}{C} (\nu \ln s^2 - \sum \nu_i \ln s_i^2)$$

where the s_i^2 are k independent sample variances based on ν_i degrees of freedom, $s^2 = \sum \nu_i s_i^2 / \nu$ is the pooled estimate based on all k sample variances, $\nu = \sum \nu_i$, and

$$C = 1 + \frac{\sum (1/\nu_i) - 1/\nu}{3(k-1)}$$

For $\nu_i \geq 5$, the distribution of β is satisfactorily approximated by a chi-square distribution with $k - 1$ degrees of freedom. In this example,

$$\begin{aligned} C &= 1 + \frac{1}{6} \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{8} - \frac{1}{26} \right) \\ &= 1.0515 \end{aligned}$$

and

$$\begin{aligned} \nu_1 &= 9 \ln s_1^2 = -5.1763 \\ \nu_2 &= 8 \ln s_2^2 = -3.6613 \\ \nu_3 &= 9 \ln s_3^2 = -3.4304 \\ \nu &= 26 \ln s^2 = -3.8599 \end{aligned}$$

so that

$$\begin{aligned} B &= \frac{26(-3.8599) - 9(-5.1763) - 8(-3.6613) - 9(-3.4304)}{1.0515} \\ &= 6.08 \end{aligned}$$

Because $\chi_{0.95}^2(2) = 5.99$, this result is just over the 0.05 level of signifi-

cance, so that the assumption of equal variances inherent in the procedures of Sections 5.1.2 and 5.1.1 should be questioned. Differences in measurement variability between laboratories in experiments of this type are not unusual. To demonstrate the multiple comparison techniques of Section 5.1.4.4, it is assumed that the use of the pooled estimate given by the mean square within laboratories is valid. The alternative approach would be to form multiple pair-wise comparisons using the methods of Section 4.3.2.2, and interpret these multiple tests based on the procedures in Section 10.2.

The study was planned to determine the possibility of bias between the shipper's and receiver's laboratories. The contrast of interest is the bias between the two laboratories

$$C_1 = (+1)\mu_1 + (-1)\mu_2 + (0)\mu_3$$

so that

$$\begin{aligned} C_1 &= \bar{x}_1 - \bar{x}_2 = \frac{1002.56}{10} - \frac{900.00}{9} \\ &= 100.256 - 100.000 = 0.256 \end{aligned}$$

Designating the mean square within laboratories by MS_w , from Equation 5.51 the estimated variance of \hat{C}_1 is

$$\begin{aligned} s_{\hat{C}_1}^2 &= MS_w \left(\frac{(1)^2}{10} + \frac{(-1)^2}{9} \right) \\ &= 0.02107 (0.1 + 0.11111) \\ &= 0.004448088 \\ s_{C_1} &= 0.06669399 \end{aligned}$$

The hypotheses being tested can be written as

$$H_0: C_1 = 0$$

$$H_A: C_1 \neq 0$$

From Equation 5.53, the statistic to test this hypothesis is

$$t = \frac{\hat{C}_1}{s_{C_1}} = \frac{0.256}{0.06669399} = 3.838$$

Because the calculated value of the statistic is larger than $t_{0.975}(26) = 2.06$ for the Student's t -distribution, the null hypothesis of no bias is rejected.

Because the referee's results apparently agree with the receiver's, a second contrast of interest is

$$C_2 = \mu_1 - \frac{\mu_2 + \mu_3}{2}$$

The hypothesis tested by this contrast is whether the shipper's laboratory is biased relative to the average of the receiver's and referee's laboratories. The hypotheses are written as

$$H_0: C_2 = 0$$

$$H_A: C_2 \neq 0$$

Again using Scheffé's test as described in Section 5.1.4.4, the test statistic is computed as follows:

$$\hat{C}_2 = \bar{x}_1 - \frac{\bar{x}_2 + \bar{x}_3}{2} = 100.256 - \frac{100.000 + 99.942}{2}$$

$$= 0.285$$

$$s_{\hat{C}_2}^2 = MS_w \left[\frac{(1)^2}{10} + \frac{(-1/2)^2}{9} + \frac{(-1/2)^2}{10} \right]$$

$$= 0.02107 [0.1527778]$$

$$= 0.00321903$$

$$s_{C_2} = 0.056736$$

$$t = \frac{0.285}{0.056736} = 5.023$$

The upper critical value for Scheffé's test is $\sqrt{(a-1)F_{1-\alpha}(a-1, N-a)}$. In this example, $a = 3$ (the number of laboratories) and $N = 29$, so that the upper critical value for $\alpha = 0.05$ is

$$\sqrt{(3-1)(3.37)} = \sqrt{2(3.37)} = \sqrt{6.74} = 2.60$$

where $F_{0.95}(2,26) = 3.37$ is obtained from Table A6. Because the calculated value of the test statistic exceeds the upper critical value, the null hypothesis is rejected.

In summary, analysis of the data indicates that there is evidence of significant differences between the three laboratories. The shipper's laboratory is biased high with respect to the receiver's laboratory and also with respect to the average of the receiver's and referee's laboratories. It appears that the main bias problem is in the shipper's laboratory because both of the other laboratories have average recovery values very close to 100.

When there is no indication of a significant difference between the laboratories (all $\mu_i = \mu$), there may still be a bias in the measurement method. This bias can be expressed as

$$\beta = \mu - \mu_0$$

where

- μ = the mean of the population of x_{ik} measurements
- μ_0 = the assigned reference value of the standard

An estimate of the bias can be computed from the equation

$$\hat{\beta} = \bar{x}_{..} - \mu_0 \quad (15.48)$$

where $\bar{x}_{..}$ is the grand average of the N standard measurements in the study.

When only a bias for the measurement method exists, the model for the measured result on the k^{th} production item made in the i^{th} laboratory can be written as

$$y_{ik} = T_k + \beta + \eta_{ij}$$

and a corrected production result can be expressed as

$$y'_{ik} = y_{ik} - \hat{\beta} \quad (15.49)$$

where

T_k = the true value of the characteristic for the k^{th} item

η_{ik} = the error effect for the measurement of the k^{th} item in the i^{th} laboratory

The variance of a corrected result can be expressed as

$$\text{Var}(y'_{ik}) = \sigma_i^2 + \sigma^2 + \sigma_0^2 \quad (15.50)$$

where

σ_i^2 = the measurement variance when measuring process materials (or items) at the i^{th} laboratory

σ^2 = the measurement variance when measuring the standard used in the study

σ_0^2 = the variance associated with the assigned value of the standard

The mean square error for an uncorrected process sample result is

$$\text{MSE}(y_{ik}) = \sigma_i^2 + \beta^2 \quad (15.51)$$

An estimate of the variance of a corrected result is

$$\text{V}\hat{\text{ar}}(y'_{ik}) = \hat{\sigma}_i^2 + \frac{\text{MS}_w}{N} + \hat{\sigma}_0^2 \quad (15.52)$$

where

$\hat{\sigma}_i^2$ = an estimate of the measurement error for process samples at the i^{th} laboratory

MS_w = the "within labs" mean square from the analysis of variance

N = the total number of standard measurements

$\hat{\sigma}_0^2$ = an estimate of the variance associated with the assigned value of the standard

Individual estimates of the measurement error variances are not obtained from the analysis of variance procedure. Procedures for obtaining individual estimates are considered in Section 15.3. If it can be demonstrated that the random error is the same for measuring process samples as it is for measuring the standard ($\sigma_i^2 = \sigma^2$), then the estimate given by Equation 15.52 could be written as

$$\text{Var}(y'_{ik}) = \text{MS}_w \left(1 + \frac{1}{N}\right) + \hat{\sigma}_0^2 \quad (15.53)$$

An unbiased estimate of the mean square error for an uncorrected value is given by

$$\text{MSE}(y_{ik}) = \hat{\sigma}_i^2 + \beta^2 - \frac{\text{MS}_w}{N} - \hat{\sigma}_0^2 \quad (15.54)$$

If it has been shown that $\sigma_i^2 = \sigma^2$, this can be written as

$$\text{M}\hat{\text{S}}\text{E}(y_{ik}) = \text{MS}_w \left(1 - \frac{1}{N}\right) + \hat{\beta}^2 - \hat{\sigma}_0^2 \quad (15.55)$$

As noted in Section 15.2.1.2, one approach to determining the need to correct measurements on process samples has been to bias-correct whenever the estimated mean square error for an uncorrected result is larger than the estimated variance of a corrected result.

When an interlaboratory study indicates significantly different biases between the laboratories, the above discussion holds for each individual laboratory if $\hat{\beta}$ is replaced by

$$\hat{\beta}_i = \bar{x}_i - \mu_0 \quad (15.56)$$

and N is replaced by n_i .

15.2.2.2 Interlaboratory Exchanges

The analytical accuracy and precision of laboratories is often evaluated by a program of interlaboratory exchanges of samples and standard-type materials. The latter may be samples from a uniform batch of typical process material or in-house standards that have been compared with standard reference materials to establish an in-house acceptance value. By mutual agreement, several laboratories will analyze the standard-type samples in round robin fashion and exchange the results for evaluation purposes. Such a program was carried out for many years under the auspices of the U. S. Department of Energy as the Safeguards Analytical Laboratory Evaluation (SALE) Program (New Brunswick Laboratory, 1983).

When samples from many different material types are distributed to several laboratories, interest in most instances is centered on an evaluation of data obtained at selected laboratories on a single material type that closely approximates a particular material routinely measured for account-

ing purposes. The data analysis and evaluation are generally carried out on one material type at a time, even though data from several standard-type materials are generated in the program. An example will be presented to demonstrate the type of evaluation that can be made using this type of data. To keep the example simple, so that the main features of the evaluation process will be clear, the data set used has the same amount of data from all laboratories. The evaluation process is equally applicable when the data sets are unequal. Such unbalanced data sets are easily analyzed with standard programs from statistical computing packages.

Example 15.11 Samples from a standard-type Pu oxide material are distributed to seven laboratories for analysis of the percent plutonium by a given analytical method. The assigned reference value for the material is not known to the participating laboratories at the time the analyses are performed. The reported percent Pu results have been coded by subtracting 86 from each result. The coded results are presented in Table 15.15.

The following quantities are calculated as a preliminary step to obtaining the sums of squares for an analysis of variance (see Section 5.1.1).

$$\begin{aligned} Q_3 &= 0.318^2 + 0.388^2 + \dots + 0.458^2 \\ &= 19.432358 \end{aligned}$$

$$\begin{aligned} Q_2 &= \frac{1}{3}(0.864^2 + 0.374^2 + \dots + 0.934^2) \\ &= \frac{1}{3}(55.433274) = 18.477758 \end{aligned}$$

$$\begin{aligned} Q_1 &= \frac{1}{9}(1.932^2 + 6.022^2 + \dots + 3.892^2) \\ &= \frac{1}{9}(162.412966) = 18.045885111 \end{aligned}$$

$$Q_0 = \frac{32.172^2}{63} = 16.429168$$

Using these quantities, the sums of squares are calculated using

$$\text{SS between laboratories} = Q_1 - Q_0$$

$$\text{SS between aliquots within laboratories} = Q_2 - Q_1$$

$$\text{SS between analyses within aliquots} = Q_3 - Q_2$$

The analysis of variance table as given in Section 5.3.2.3 is presented in

TABLE 15.15
Plutonium Concentration Results (Coded) from Seven Laboratories

| Laboratory
Aliquot | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|-------|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| Analyses | 0.318 | 0.048 | 0.128 | 0.580 | 0.730 | 0.680 | 0.587 | 0.507 | 0.587 | 0.690 | 0.662 | 0.710 |
| | 0.388 | 0.228 | 0.398 | 0.660 | 0.890 | 0.110 | 0.567 | 0.457 | 0.587 | 0.683 | 0.546 | 0.865 |
| | 0.158 | 0.098 | 0.168 | 0.610 | 0.872 | 0.890 | 0.547 | 0.467 | 0.487 | 0.793 | 0.596 | 0.932 |
| Sums | 0.864 | 0.374 | 0.694 | 1.850 | 2.492 | 1.680 | 1.701 | 1.431 | 1.661 | 2.166 | 1.804 | 2.507 |
| Laboratory totals | | | 1.932 | | | 6.022 | | | 4.793 | | | 6.477 |
| Laboratory
Aliquot | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 7 | | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | | | |
| Analyses | 0.570 | 0.867 | 0.503 | 0.380 | 0.400 | 0.440 | 0.478 | 0.418 | 0.158 | | | |
| | 0.542 | 0.390 | 0.242 | 0.170 | 0.610 | 0.380 | 0.698 | 0.548 | 0.318 | | | |
| | 0.828 | 0.782 | 0.622 | 0.510 | 0.370 | 0.450 | 0.328 | 0.488 | 0.458 | | | |
| Sums | 1.940 | 2.039 | 1.367 | 1.060 | 1.380 | 1.270 | 1.504 | 1.454 | 0.934 | | | |
| Laboratory totals | | | 5.346 | | | 3.710 | | | 3.892 | | | |
| Grand total | | | | | | | | | | 32.172 | | |

Table 15.16. In this case, σ_β^2 is the variance between aliquots and the α_i are fixed effects associated with the laboratories in the study.

TABLE 15.16
Analysis of Variance for Interlaboratory Comparisons
of Plutonium Concentrations

| Source | df | SS | MS | E(MS) | F |
|--------------|----|-----------|-----------|---|-------|
| Laboratories | 6 | 1.6167171 | 0.2694529 | $\sigma^2 + 3\sigma_\beta^2 + 3(3) \Sigma \alpha_i^2 / 2$ | 8.735 |
| Aliquots | 14 | 0.4318729 | 0.0308481 | $\sigma^2 + 3\sigma_\beta^2$ | 1.357 |
| Analyses | 42 | 0.9546000 | 0.0227286 | σ^2 | |

The estimates of σ^2 and σ_β^2 are

$$\hat{\sigma}^2 = 0.02273$$

$$\begin{aligned} \hat{\sigma}_\beta^2 &= \frac{1}{3}(0.0308481 - 0.0227286) \\ &= 0.00271 \end{aligned}$$

When, as in this example, the estimated value of σ_β^2 is relatively small and the F-ratio between the mean squares for aliquots and analyses is not significant [$F_{0.95}(14,42) = 1.93$], there are two possible approaches to the testing of differences between laboratories. The first is to base the tests on a variance estimate obtained by pooling the SS(analyses) and SS(aliquots) and the df(analyses) and df(aliquots) to obtain a new estimate of MS(analyses) under the assumption that $\sigma_\beta^2 = 0$. For this example a pooled mean square, MS_p , of $(0.4318729 + 0.9546)/(14 + 42) = 0.0247584$ is obtained. Using this approach, the F-ratio of 10.883 between the mean square for laboratories and the pooled mean square is used to test the hypothesis of no differences between laboratories. The tabulated value of $F_{0.95}(6,56)$ is 2.266, so there is evidence of significant differences between laboratories. The alternative approach is to use the variance estimate based on the mean square for aliquots to test differences between laboratories regardless of whether or not the first test is significant.

The arguments for the two approaches are similar to those involved in the decision to correct for bias. The pooled estimate will usually be based on a substantially greater number of degrees of freedom, and hence, may be more reliable. On the other hand, if σ_β^2 is not actually zero, the pooled estimate will underestimate the true contribution of the aliquoting and

analysis to the variability between laboratories. A useful rule of thumb has been to pool whenever the F-ratio between the two estimates to be pooled is both nonsignificant at the 0.05 level and less than 2. Alternatively, the approximate test given in Section 5.3.1.1 can be used. In the remaining parts of this example, the approach using the pooled mean square will be used for further study.

Because evidence exists that the laboratories are different, the question arises concerning which laboratories are similar and which ones are different. Duncan's multiple range test can be applied to help identify which of the population means differ. As in Section 5.1.4.3, the laboratory averages are arranged in ascending order and values of R_1' through R_7 are calculated and used to compare and test the observed differences between means.

| Laboratory | 1 | 6 | 7 | 3 | 5 | 2 | 4 |
|------------|--------|--------|--------|--------|--------|--------|--------|
| Average | 0.2147 | 0.4122 | 0.4324 | 0.5326 | 0.5940 | 0.6691 | 0.7197 |

Using $\alpha = 0.05$, the q values needed for calculating the R_p values were obtained through simple linear interpolation from the tables for Duncan's multiple range test. The values of q and the corresponding R_p values for $p = 2$ to 7 are also listed as follows:

| | |
|------------------|----------------|
| $q(7,56) = 3.25$ | $R_7 = 0.1705$ |
| $q(6,56) = 3.20$ | $R_6 = 0.1678$ |
| $q(5,56) = 3.15$ | $R_5 = 0.1652$ |
| $q(4,56) = 3.08$ | $R_4 = 0.1615$ |
| $q(3,56) = 2.99$ | $R_3 = 0.1568$ |
| $q(2,56) = 2.84$ | $R_2 = 0.1490$ |

The R_p values were calculated using $S_{\bar{y}} = \sqrt{.0247584/9} = 0.0524493$ obtained from the pooled estimate of σ^2 . The results for all of the comparisons between means are summarized in the following table.

| Laboratory | 1 | 6 | 7 | 3 | 5 | 2 | 4 |
|------------|----|------|------|------|------|------|------|
| 1 | ok | diff | diff | diff | diff | diff | diff |
| 6 | | ok | ok | ok | diff | diff | diff |
| 7 | | | ok | ok | diff | diff | diff |
| 3 | | | | ok | ok | ok | diff |
| 5 | | | | | ok | ok | ok |
| 2 | | | | | | ok | ok |
| 4 | | | | | | | ok |

As frequently happens with this type of data evaluation, the results are not definitive. It is not clear whether Laboratory 3 belongs with Laboratories 6 and 7 or with Laboratories 5 and 2, nor whether Laboratory 5 belongs to the highest or next highest group.

This analysis indicates that all of the laboratories should examine their use of this analytical method in an effort to detect a possible bias. If Laboratory 4 were making the shipper's measurements and Laboratory 1 were making the receiver's measurements, significant shipper-receiver differences would be highly probable.

15.3 ESTIMATES FROM REPEATED MEASUREMENTS

The previous sections have been concerned with the estimation of error parameters from the monitoring of direct measurements against standards or the determination of calibration functions for indirect measurement procedures. In addition to these monitoring and calibration activities based on external standards, routine control or verification procedures or inter-laboratory comparisons often involve repeated measurements from which error parameters can be estimated. When nonstandard (production) materials are used in such studies, estimates and tests for bias are not possible. However, the error variances are frequently more realistic when obtained from measurement data on production items than those obtained from measurements on standards. There are two main reasons for this. First, it is often difficult or impossible to obtain standards that are typical of the production items in all important respects such as composition, impurities, interferences, geometry, etc. Second, if it is recognized that the measurement is being made on a standard, special or an excessive amount of care is often exercised.

15.3.1 Data from a Single Source

When repeated measurements are made on the same or similar production items, it is possible to obtain a direct estimate of the error variance. Care must be taken to recognize which error sources are replicated in such measurements and which are constant for all data in the set. Also, the time period covered by the repeated measurements should be one in which the calibration and control of the measurement method has been stable.

The data obtained where a single sealed can of UO_2 powder is repeatedly weighed on a given scale at random intervals over a 1-month time period provide a simple example. If it is known from past experience that

the calibration of the scale can be expected to remain valid for longer than 1 month, and that the contents of the can remain constant during the month, then such data can be used to calculate an estimate of the random error variance.

Example 15.12 The following are 24 weight measurement results (in kg) for one can:

| Measurement | Weight (kg) | Measurement | Weight (kg) |
|-------------|-------------|-------------|-------------|
| 1 | 22.038 | 13 | 22.036 |
| 2 | 22.041 | 14 | 22.037 |
| 3 | 22.033 | 15 | 22.038 |
| 4 | 22.048 | 16 | 22.029 |
| 5 | 22.038 | 17 | 22.029 |
| 6 | 22.043 | 18 | 22.038 |
| 7 | 22.029 | 19 | 22.044 |
| 8 | 22.047 | 20 | 22.037 |
| 9 | 22.047 | 21 | 22.029 |
| 10 | 22.033 | 22 | 22.047 |
| 11 | 22.031 | 23 | 22.049 |
| 12 | 22.038 | 24 | 22.032 |

The sample mean, variance, and standard deviation are

$$\bar{x} = 22.0380$$

$$s^2 = 0.000043435$$

$$s = 0.00659$$

If a long history using this particular make and model of scale in similar applications in the plant has shown that the expected variance is $\sigma_0^2 = (0.0050)^2 = 0.000025$, then the hypothesis that the variance for this scale during this period was equal to the historical value σ_0^2 can be tested. The case where $\sigma^2 > \sigma_0^2$ would have practical importance, because it would indicate possible need for maintenance on, or replacement of, the scale. Using a chi-square test as described in Section 4.2.3 at the 0.05 level of significance, the test statistic is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = 23 \frac{(0.000043435)}{0.000025} = 39.960$$

Because $\chi_{0.95}^2(23) = 35.2$, there is reason to reject the null hypothesis that the population variance for this scale is equal to the historical or standard value. If a standard deviation equal to σ_0 is required when using this scale, then replacement or repair is indicated.

The use of a single can of powder for such a large number of repeated measurements could lead to recognition of the can by the person doing the weighing. The 24 measurements might have been more effectively made on several different cans of powder submitted at random over the month. Instead of the data in the above example, the following 26 measurement results might have been observed:

| Can 1 | Can 2 | Can 3 | Can 4 | Can 5 |
|--------|--------|--------|--------|--------|
| 22.038 | 22.616 | 21.418 | 19.811 | 24.095 |
| 22.041 | 22.615 | 21.425 | 19.825 | 24.120 |
| 22.033 | 22.617 | 21.414 | 19.808 | 24.096 |
| 22.048 | 22.608 | | 19.802 | 24.105 |
| | 22.603 | | 19.795 | 24.105 |
| | 22.610 | | 19.799 | 24.118 |
| | | | | 24.113 |

The scatter in the n_i repeated measurements for each can is combined to estimate the random error variance from these data. A one-way analysis of variance can be used, which in addition to providing the random error variance estimate allows an estimate of the variance σ_α^2 due to the actual weight difference between cans. If the five cans were randomly selected, then such an estimate can be obtained from a one-way analysis of variance. Using the methods in Section 5.2.1 for random effects, the following ANOVA table was obtained:

| Source | df | SS | MS | E(MS) |
|--------------|----|------------|-------------|---------------------------------|
| Between cans | 4 | 62.6954338 | 15.67385845 | $\sigma^2 + n_0\sigma_\alpha^2$ |
| Within cans | 21 | 0.0015006 | 0.000071457 | σ^2 |
| Total | 25 | | | |

where

$$\begin{aligned} n_0 &= \frac{1}{(5-1)} (26 - \sum n_i^2/26) = \frac{1}{4} (26 - 5.615384615) \\ &= 5.0962 \end{aligned}$$

From this data set, the estimated random error variance due to measurement is $0.000071457 \text{ (kg)}^2$. A chi-square test of this residual variance estimate based on 21 degrees of freedom against the hypothetical variance $\sigma_0^2 = 0.000025$ leads to the test statistic $\chi^2 = 60.024$, which exceeds $\chi_{0.95}^2(21) = 32.7$, again indicating need for possible replacement or repair of the scale.

15.3.2 Data from Two Sources

The simplest treatment of data from two sources occurs when it can be assumed that the two measurement methods (or the two parties) have the same random measurement standard deviation. Let

- x_{1k} = measurement result for item k using measurement method 1
- x_{2k} = measurement result for item k using measurement method 2
- n = number of items independently measured using both methods
- $d_k = x_{1k} - x_{2k}$
- σ^2 = the measurement variance for both methods

There are two possible estimators for σ^2 depending on whether or not the two measurement methods are biased relative to each other. Under the assumption of no relative bias between the methods, the estimator

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^n d_k^2}{2n} \quad (15.57)$$

is the appropriate one to use. Under the assumption that there is a bias between the methods, the estimator

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^n d_k^2 - \left(\sum_{k=1}^n d_k \right)^2 / n}{2(n-1)} \quad (15.58)$$

should be used.

Example 15.13 A sample of uniform size is drawn from each of several containers of nitrate solution and analyzed for percent plutonium content using two different analytical methods. Historical quality control data have routinely confirmed that the two particular methods used in the study have the same random measurement variance. Sample results for 20 containers are given below. Two questions need to be answered: Is the bias between the two methods equal to zero? What is the current estimate of the common random standard deviation?

| Container | x_{1k} | x_{2k} | x_{dk} |
|-----------|----------|----------|----------|
| 1 | 13.11 | 13.00 | 0.11 |
| 2 | 15.14 | 14.90 | 0.24 |
| 3 | 13.22 | 13.01 | 0.21 |
| 4 | 13.67 | 13.65 | 0.02 |
| 5 | 10.48 | 10.61 | -0.13 |
| 6 | 15.37 | 15.11 | 0.26 |
| 7 | 12.37 | 12.40 | -0.03 |
| 8 | 12.50 | 12.63 | -0.13 |
| 9 | 11.46 | 11.71 | -0.25 |
| 10 | 14.28 | 14.21 | 0.07 |
| 11 | 13.26 | 13.01 | 0.25 |
| 12 | 11.00 | 11.06 | -0.06 |
| 13 | 12.74 | 12.75 | -0.01 |
| 14 | 13.69 | 13.69 | 0.00 |
| 15 | 10.43 | 10.40 | 0.03 |
| 16 | 11.38 | 11.30 | 0.08 |
| 17 | 12.26 | 12.27 | -0.01 |
| 18 | 12.89 | 12.70 | 0.19 |
| 19 | 13.33 | 13.30 | 0.03 |
| 20 | 11.88 | 11.90 | -0.02 |

From these data,

$$\sum d_i = 0.85$$

$$\sum d_k^2 = 0.3949$$

$$\bar{d} = 0.0425$$

$$s_d^2 = \frac{0.3949 - (0.85)^2/20}{(19)}$$

$$= 0.018882894$$

If there is no bias, σ can be estimated as

$$\hat{\sigma}^2 = \frac{\sum d_k^2}{2(20)} = \frac{0.3949}{40} = 0.0098725$$

$$\hat{\sigma} = 0.09936$$

If the absence of bias cannot be assumed, then the appropriate estimate of σ is

$$\hat{\sigma}^2 = s_d^2/2 = \frac{0.018882894}{2} = 0.0094414$$

$$\hat{\sigma} = 0.09717$$

To test the hypothesis that there is no systematic difference between the two measurement methods, the approach used in Section 4.2.2 can be applied. The hypotheses are:

$$H_0: \mu_d = 0$$

$$H_A: \mu_d \neq 0$$

If the test is made at the 0.05 level of significance, the critical values -2.093 and $+2.093$ are obtained from a table of Student's *t*-distribution for 19 degrees of freedom. The calculated value for the test statistic *t* for this example is

$$t = \frac{\bar{d}}{\sigma_d/\sqrt{n}} = \frac{0.0425}{\frac{0.137415}{\sqrt{20}}} = 1.383$$

so the null hypothesis is not rejected. There is no evidence that the two methods are biased with respect to each other.

So far in this section, it has been assumed that the error variances for the two parties or methods are the same. Although paired data occur naturally in both shipper–receiver comparisons and comparisons of measurement results on the same items during inspections, the measurement methods used by the two parties may be quite different and the assumption of equal variances is often not appropriate.

A general model for this situation is given by

$$\begin{aligned}x_{1k} &= \mu + \theta_k + \epsilon_{1k} \\x_{2k} &= \mu + \Delta + \theta_k + \epsilon_{2k}\end{aligned}\tag{15.59}$$

where μ is the overall mean of the n items and Δ is a possible bias between the paired results. The ϵ_{1k} , ϵ_{2k} , and θ_k are assumed to be random error effects associated with the two measurement methods and the variation in the true content of the items measured. The model is similar to that for a $2 \times n$ random effects analysis of variance (Section 5.4.3) except that the variances of the ϵ_{1k} and ϵ_{2k} are not assumed to be equal. Assume the random effects to be independent with expected value zero, and let $E(\epsilon_{1k}^2) = \sigma_1^2$, $E(\epsilon_{2k}^2) = \sigma_2^2$, and $E(\theta_k^2) = \sigma_\theta^2$ (commonly called the product variance). Then if

s_1^2 = sample variance for the x_{1k}

s_2^2 = sample variance for the x_{2k}

s_{12} = sample covariance for the x_{1k} x_{2k} pairs

it is easily shown (Section 2.7) that

$$E(s_1^2) = \sigma_\theta^2 + \sigma_1^2$$

$$E(s_2^2) = \sigma_\theta^2 + \sigma_2^2$$

$$E(s_{12}) = \sigma_\theta^2$$

$$E(s_\theta^2) = \sigma_1^2 + \sigma_2^2$$

The last value can be derived directly from the model or computed from the fact that $E(s_1^2) + E(s_2^2) - 2E(s_{12}) = E(s_\theta^2)$. As noted in Section 4.4, the covariance (or correlation) between the measured values is a direct reflection of the common variability in the two measurements, which in this case is simply the variability in the items being measured.

Using the first three equations and replacing expected values by observed sample variances, the following estimates are obtained:

$$\hat{\sigma}_1^2 = s_1^2 - s_{12}$$

$$\hat{\sigma}_2^2 = s_2^2 - s_{12}$$

$$\hat{\sigma}_\theta^2 = s_{12}\tag{15.60}$$

Equivalently, the three estimates can be expressed as

$$\begin{aligned}\hat{\sigma}_1^2 &= (s_1^2 + s_d^2 - s_2^2)/2 \\ \hat{\sigma}_2^2 &= (s_2^2 + s_d^2 - s_1^2)/2 \\ \hat{\sigma}_\theta^2 &= (s_1^2 + s_2^2 - s_d^2)/2\end{aligned}\quad (15.61)$$

This form reflects the method of generalizing to more than two measurement methods considered in the next section.

This approach to the estimation of unequal error variances, along with the methods to be considered in the next section, was proposed by F. E. Grubbs (1948). Assuming the measurement errors to be normally distributed, the variances of the estimates of the measurement errors are given by

$$\text{Var}(\hat{\sigma}_1^2) = \frac{2\sigma_1^4 + \sigma_\theta^2 \sigma_1^2 + \sigma_\theta^2 \sigma_2^2 + \sigma_1^2 \sigma_2^2}{n-1} \quad (15.62)$$

$$\text{Var}(\hat{\sigma}_2^2) = \frac{2\sigma_2^4 + \sigma_\theta^2 \sigma_1^2 + \sigma_\theta^2 \sigma_2^2 + \sigma_1^2 \sigma_2^2}{n-1} \quad (15.63)$$

When the number of measurements is small and either the other measurement variance or the product variance, or both, are large relative to the measurement variance being estimated, the estimate $\hat{\sigma}_1$ will be highly variable because $\text{SE}(\hat{\sigma}_1^2) = \sqrt{\text{Var}(\hat{\sigma}_1^2)}$ will be large compared to σ_1^2 . This can result in a high frequency of negative variance estimates. On the other hand, if the product variance is small compared to the measurement variances, the error in direct estimates of the measurement variances based on the sample variances s_1^2 and s_2^2 , without any correction for a common effect, is small. Methods have been proposed (Jaech, 1981) that will always lead to positive estimates, but these do not improve their basic quality.

Example 15.14 This example presents a situation where the Grubbs method can be used to obtain reasonable estimates of the measurement error variances. Eighteen containers of solution were analyzed for percent uranium using two different direct chemical methods which are expected to have different error variances. The results are listed below:

| Container | x_{1k} | x_{2k} | d_k |
|-----------|----------|----------|-------|
| 1 | 75.44 | 73.96 | 1.48 |
| 2 | 77.46 | 75.98 | 1.48 |
| 3 | 72.22 | 74.15 | -1.93 |
| 4 | 75.85 | 75.98 | -0.13 |
| 5 | 74.28 | 77.44 | -3.16 |
| 6 | 76.82 | 77.61 | -0.79 |
| 7 | 74.24 | 70.30 | 3.94 |
| 8 | 77.87 | 80.27 | -2.40 |
| 9 | 75.32 | 78.75 | -3.43 |
| 10 | 76.17 | 73.70 | 2.47 |
| 11 | 73.21 | 77.93 | -4.72 |
| 12 | 75.65 | 74.70 | 0.95 |
| 13 | 76.93 | 73.38 | 3.55 |
| 14 | 72.36 | 73.67 | -1.31 |
| 15 | 79.15 | 76.01 | 3.14 |
| 16 | 75.90 | 77.01 | -1.11 |
| 17 | 77.03 | 76.71 | 0.32 |
| 18 | 75.90 | 73.71 | 2.19 |

For this example

$$s_1^2 = 3.433038$$

$$s_2^2 = 5.751697$$

$$s_{12} = 1.317212$$

and hence

$$\hat{\sigma}_1^2 = 2.1158 \quad \text{or} \quad \hat{\sigma}_1 = 1.455$$

$$\hat{\sigma}_2^2 = 4.4345 \quad \text{or} \quad \hat{\sigma}_2 = 2.106$$

Both variance estimates turn out to be positive for this data set. However, if the paired measurements for the percent uranium on the 18 containers were repeated several times, the expected variability in the estimates obtained from the different data sets would be quite large.

It may be desirable to test the hypothesis of no difference between the two measurement variances, i.e., $H_0: \sigma_1^2 = \sigma_2^2$. An attempt to perform an

F-test, as presented in Section 4.3.3, is not appropriate because the estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ do not have chi-square distributions and hence, their ratio does not have an F-distribution. An alternate approach has been proposed (Maloney and Rastagi, 1971) which does lead to a valid test of the hypothesis. If the additional variable $u_k = x_{1k} + x_{2k}$ is introduced, the following quantities can be calculated:

$$\begin{aligned} s_u^2 &= s_1^2 + s_2^2 + 2s_{12} \\ s_d^2 &= s_1^2 + s_2^2 - 2s_{12} \\ s_{ud} &= s_1^2 - s_2^2 \end{aligned} \quad (15.64)$$

An estimate of the correlation coefficient of u and d is

$$r_{ud} = \frac{s_{ud}}{s_u s_d} \quad (15.65)$$

Because the expected value of the numerator is $\sigma_1^2 - \sigma_2^2$, a test of the hypothesis that the population correlation coefficient of u and d is zero is equivalent to the hypothesis that $\sigma_1^2 = \sigma_2^2$. By using the test statistic

$$t = \frac{r_{ud}}{\sqrt{1 - r_{ud}^2}} \sqrt{(n-2)} \quad (15.66)$$

and comparing the calculated result with the critical values of Student's t -distribution with $(n - 2)$ degrees of freedom at the α significance level, a test of the desired hypothesis can be performed.

Example 15.15 Although the hypothesis $\sigma_1^2 = \sigma_2^2$ is clearly not defensible in an exact sense for two methods known to be different, it might be desirable to determine whether the data of Example 15.14 support the hypothesis of approximately equal variances. To do this, calculate

$$\begin{aligned} s_u^2 &= 3.433038 + 5.751697 + 2(1.317212) \\ &= 11.819159 \end{aligned}$$

$$\begin{aligned} s_d^2 &= 3.433038 + 5.751697 - 2(1.317212) \\ &= 6.550311 \end{aligned}$$

$$\begin{aligned} s_{ud} &= 3.433038 - 5.751697 \\ &= -2.318659 \end{aligned}$$

from which

$$\begin{aligned} r_{ud} &= \frac{-2.318659}{\sqrt{(6.550311)(11.819159)}} \\ &= -0.26352 \end{aligned}$$

and

$$\begin{aligned} t &= \frac{-0.26352\sqrt{16}}{\sqrt{1 - (-0.26352)^2}} \\ &= -1.093 \end{aligned}$$

Because the critical values of Student's t for a two-sided test at the 0.05 level of significance with 16 degrees of freedom are ± 2.120 , the hypothesis of equal variance cannot be rejected on the basis of these measurements. If the variances are assumed to be equal, then the estimate of the common variance is given by

$$\begin{aligned} \hat{\sigma}^2 &= s_d^2/2 = 3.2752 \\ \hat{\sigma} &= 1.810 \end{aligned}$$

This estimate is exactly the same as the residual variance estimate that would be obtained if these data were analyzed by the methods of Chapter 5 using a $2 \times n$ two-way analysis of variance. Also, $s_{12} = (s_u^2 - s_d^2)/4$ is precisely the estimate of the variance of the item effects obtained by differencing the expected mean squares for a two-way random effects ANOVA model because $s_u^2/2$ is exactly the between-items mean square in the $2 \times n$ case. The validity of the between-items estimate does not depend on equality of the measurement error variances.

15.3.3 Data from Three or More Sources

When results are available from more than two methods or laboratories on the same set of n items, it is no longer necessary to involve the product variance to obtain estimates of the error variances for individual laboratories or methods. The following sections describe two closely related

approaches. The first was given by Grubbs (1948) in the same article in which the method for two sources was introduced. The second is contained in an unpublished report by Stewart (1980) which was primarily concerned with outlier detection and elimination prior to the use of Grubb's procedures.

The basic data consist of the results x_{ik} obtained by $i=1, \dots, m$ instruments (laboratories or measurement methods) on each of $k=1, \dots, n$ items to be measured. The data can be considered as the results of a randomized block experiment where the items being measured are blocks and the methods used for measurement, or the laboratories performing the measurements, are the treatments. The analysis differs from that given in Sections 5.4.3 and 7.2.2 only in that the measurement error effects for each treatment are assumed to be from different populations.

15.3.3.1 Grubbs Approach

Let $d_{ijk} = x_{ik} - x_{jk}$ be the difference between the measurements of the i^{th} and j^{th} laboratories on the k^{th} item. Then the average difference for the n items is

$$\bar{d}_{ij} = \sum_{k=1}^n d_{ijk}/n = \bar{x}_i - \bar{x}_j \quad (15.67)$$

The sample variance of the differences is

$$\begin{aligned} V_{ij} &= s_{d_{ij}}^2 = \sum_{k=1}^n (d_{ijk} - \bar{d}_{ij})^2 / (n-1) \\ &= \sum_{k=1}^n (x_{ik} - x_{jk} - \bar{x}_i + \bar{x}_j)^2 / (n-1) \end{aligned} \quad (15.68)$$

The general model for the results is assumed to be

$$x_{ik} = \mu + \beta_i + \theta_k + \epsilon_{ik}$$

where

- μ = overall mean
- β_i = instrument effects
- θ_k = item effects
- ϵ_{ik} = individual measurement errors

If it is assumed that $E(\epsilon_{ik}) = 0$ and $E(\epsilon_{ik}^2) = \sigma_1^2$ for repeated measurements on the same item by a given instrument, then

$$E(V_{ij}) = \sigma_i^2 + \sigma_j^2$$

If all $m(m - 1)/2$ possible paired differences of the m instruments with $i < j$ are considered, then exactly $m - 1$ of the V_{ij} will have a specific σ_i^2 in their expected values, and these $m - 1$ expected values will contain each of the other error variances σ_j^2 , $j \neq i$, exactly once. The expected values of the remaining $(m - 1)(m - 2)/2$ sample variances of the differences contain each σ_j^2 , $j \neq i$, exactly $m - 2$ times. Hence, an estimate of σ_i^2 is given by

$$\hat{\sigma}_i^2 = \sum_{r=1}^m \sum_{s=r+1}^m a_{rs} V_{rs} \quad (15.69)$$

where $a_{rs} = (m - 1)^{-1}$ if either r or s is equal to i and $a_{rs} = -[(m - 1)(m - 2)]^{-1}$ if neither r nor s is equal to i . For example, if $m = 4$

$$\hat{\sigma}_2^2 = \frac{V_{12} + V_{23} + V_{24}}{3} - \frac{V_{13} + V_{14} + V_{34}}{(3)(2)}$$

Since $V_{ii} = 0$ and $V_{ij} = V_{ji}$, $S_i = \sum_{j=1}^m V_{ij}$ is equal to the sum of terms in $\hat{\sigma}_i^2$ that have the coefficient $(m - 1)^{-1}$. Let $S = \sum_{i=1}^m S_i$. Then $S/2 - S_i$ is equal to the sum of the terms in $\hat{\sigma}_i^2$ for which the coefficients are $-1/[(m - 1)(m - 2)]$. Multiplying the sums by their appropriate coefficients and adding yields

$$\begin{aligned} \hat{\sigma}_i^2 &= \frac{S_i}{m-1} - \frac{(S/2 - S_i)}{(m-1)(m-2)} \\ &= \left[S_i = \frac{S}{2(m-1)} \right] / (m-2) \end{aligned} \quad (15.70)$$

For $m = 3$, it can be shown that these estimates are maximum likelihood estimates if the measurement error effects are assumed to be normally distributed (Jaech, 1983).

Example 15.16 Aliquots from $n = 8$ product batches were analyzed for plutonium by $m = 4$ laboratories. The results (g/l) and the sample means for each item, each instrument, and overall are given in Table

15.17. The differences between the laboratory means and the overall mean are used in a later example.

TABLE 15.17
Instrument

| Item | 1 | 2 | 3 | 4 | \bar{x}_k |
|----------------------------|-----------|-----------|-----------|-----------|----------------------------|
| 1 | 163.92 | 167.04 | 168.10 | 170.36 | 167.3550 |
| 2 | 163.63 | 163.43 | 162.74 | 164.10 | 163.4750 |
| 3 | 187.07 | 186.64 | 189.03 | 187.43 | 187.5425 |
| 4 | 182.45 | 185.80 | 186.90 | 186.41 | 185.3900 |
| 5 | 184.25 | 185.94 | 184.27 | 184.27 | 184.6825 |
| 6 | 171.39 | 172.03 | 173.35 | 173.63 | 172.6000 |
| 7 | 183.59 | 180.88 | 180.92 | 179.70 | 181.2725 |
| 8 | 159.07 | 160.69 | 159.60 | 159.84 | 159.8000 |
| \bar{x}_i | 174.42125 | 175.30625 | 175.61375 | 175.71750 | $\bar{x}_{..} = 175.26469$ |
| $\bar{x}_i - \bar{x}_{..}$ | -0.84344 | 0.04156 | 0.34906 | 0.45281 | |

The six possible differences between laboratories for each item and their sample variances for the eight items measured are shown in Table 15.18.

TABLE 15.18
Differences

| Item | 1 and 2 | 1 and 3 | 1 and 4 | 2 and 3 | 2 and 4 | 3 and 4 |
|----------|----------|----------|----------|----------|----------|----------|
| 1 | -3.12 | -4.18 | -6.44 | -1.06 | -3.32 | -2.26 |
| 2 | 0.20 | 0.89 | 0.47 | 0.69 | -0.67 | -1.36 |
| 3 | 0.43 | -1.96 | -0.36 | -2.39 | -0.79 | 1.60 |
| 4 | -3.35 | -4.45 | -3.96 | -1.10 | -0.61 | 0.49 |
| 5 | -1.69 | -0.02 | -0.02 | 1.67 | 1.67 | 0.00 |
| 6 | -0.64 | -1.96 | -2.24 | -1.32 | -1.60 | -0.28 |
| 7 | 2.71 | 2.67 | 3.89 | -0.04 | 1.18 | 1.22 |
| 8 | -1.62 | -0.53 | -0.77 | 1.09 | 0.85 | -0.24 |
| Sums | -7.08 | -9.54 | -10.37 | -2.46 | -3.29 | -0.83 |
| V_{ij} | 4.021457 | 5.969136 | 9.258084 | 1.926621 | 2.654041 | 1.613655 |

The sums of the variances are given by

$$S_1 = 4.021457 + 5.969136 + 9.258084 = 19.248677$$

$$S_2 = 4.021457 + 1.926621 + 2.654041 = 8.602119$$

$$S_3 = 5.969136 + 1.926621 + 1.613655 = 9.509412$$

$$S_4 = 9.258084 + 2.654041 + 1.613655 = 13.525780$$

$$S = \sum S_i = 50.885988$$

For $m = 4$, $\hat{\sigma}_1^2 = (S_1 - S/6)/2$ and hence,

$$\hat{\sigma}_1^2 = 5.383839$$

$$\hat{\sigma}_2^2 = 0.060561$$

$$\hat{\sigma}_3^2 = 0.514207$$

$$\hat{\sigma}_4^2 = 2.522391$$

The relative uncertainties associated with these estimates are large considering that only eight items and four instruments are involved, so the numbers of significant figures that are shown in the estimates are not warranted in practice. They are given here only to facilitate comparison with the results obtained by the alternate method developed in the next section.

15.3.3.2 Analysis of Variance Approach

The usual two-way random effects analysis of variance procedure (Section 5.4.3) can be used to fit the linear model of the previous section to the results x_{ik} . The residuals from this fit are given by

$$r_{ik} = x_{ik} - \bar{x}_i - \bar{x}_k + \bar{x}_.$$

Defining the measurement error variances for each instrument as in the previous section, it can be shown that

$$E(r_{ik}^2) = \frac{(n-1)}{m^2n} \left[(m-1)^2 \sigma_i^2 + \sum_{j \neq i}^m \sigma_j^2 \right]$$

and hence,

$$\begin{aligned} E\left[\sum_{k=1}^n \frac{r_{ik}^2}{n-1}\right] &= \sum_{k=1}^n \frac{E(r_{ik}^2)}{n-1} \\ &= \frac{1}{m^2} \left[(m-1)^2 \sigma_i^2 + \sum_{j \neq i}^m \sigma_j^2 \right] \end{aligned}$$

Note that if $\sigma_i = \sigma$ for all i , the right-hand quantity reduces to $\left(\frac{m-1}{m}\right)\sigma^2$ as expected. This set of equations can be solved for estimates of the σ_i^2 in a similar fashion to that used in the previous section, obtaining

$$\hat{\sigma}_i^2 = \frac{mR_i - R}{m-2}$$

where

$$R_i = \frac{\sum_{k=1}^n r_{ik}^2}{n-1}$$

and

$$R = \frac{\sum R_i}{m-1}$$

With some straightforward but tedious algebraic manipulation, it can be shown that these estimates are identical to those of the preceding section.

Example 15.17 The residuals for the data of Example 15.16 are

| | 1 | 2 | 3 | 4 |
|---|---------|---------|---------|---------|
| 1 | -2.5916 | -0.3566 | 0.3959 | 2.5522 |
| 2 | 0.9984 | -0.0866 | -1.0841 | 0.1722 |
| 3 | 0.3709 | -0.9441 | 1.1384 | -0.5653 |
| 4 | -2.0966 | 0.3684 | 1.1609 | 0.5672 |
| 5 | 0.4109 | 1.2159 | -0.7616 | -0.8653 |
| 6 | -0.3666 | -0.6116 | 0.4009 | 0.5772 |
| 7 | 3.1609 | -0.4341 | -0.7016 | -2.0253 |
| 8 | 0.1134 | 0.8484 | -0.5491 | -0.4128 |

The following intermediate quantities are computed for each laboratory:

| Laboratory | $\sum_{k=1}^8 r_{ik}^2$ | R_i |
|------------|-------------------------|----------|
| 1 | 22.553874 | 3.221982 |
| 2 | 3.922399 | 0.560343 |
| 3 | 5.510162 | 0.787166 |
| 4 | 12.538805 | 1.791258 |

From these results, $R = \Sigma R_i/3 = 2.120250$, and because for $m = 4$, $\hat{\sigma}_i^2 = (4R_i - R)/2$,

$$\hat{\sigma}_1^2 = 5.383839$$

$$\hat{\sigma}_2^2 = 0.060561$$

$$\hat{\sigma}_3^2 = 0.514207$$

$$\hat{\sigma}_4^2 = 2.522391$$

as in the previous section.

In the situation where an unprogrammed desk calculator is used, the difference method is probably easier than the residual method when m is small because differences are easier to calculate than residuals. However, because $m(m-1)/2$ sets of differences and only m sets of residuals have to be calculated, the difference in sets required, $m(m-1)/2 - m = m(m-3)/2$, becomes large as m increases. Thus, the residual method may be easier for larger m . Also, this approach can be applied directly to the residuals calculated in an analysis of variance and it can be appended easily to the computer programs normally available. Finally, the residuals can be easily and automatically screened for outliers using the regression residual approach given in Section 10.5.3.

If the value of n used in the computation of the V_{ij} and the R_i is adjusted for missing differences or residuals, missing values do not affect either method when the residuals are estimated through appropriate

nonorthogonal procedures. The essential features are that the expected values involve only linear combinations of the measurement variances and that the number of quantities for which independent combinations are available is sufficient to estimate individual variance components.

REFERENCES

- Bennett, C. A., and N. L. Franklin, 1954, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons Inc., New York.
- Eisenhart, C., 1939, The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research, *Ann. Math. Stat.*, **10**: 162-186.
- Goldman, A., L. G. Anderson, and J. Weber, 1983, Cumulative Error Models for the Tank Calibration Problem, *Nucl. Mater. Manage., Proc. Issue*, 24th Annual Meeting, July 10-13, 1983, **XII**: 187-194.
- Grubbs, F. E., 1948, Estimating Precision of Measurement Instruments and Product Variability, *J. Am. Stat. Assoc.*, **43**: 243-264.
- IAEA, 1980, *IAEA Safeguards Technical Manual*, TEC DOC-216, Vol. 3, Part F, International Atomic Energy Agency, Vienna.
- Jaech, J. L., 1964, A Note on the Equivalence of Two Methods of Fitting a Straight Line Through Cumulative Data, *J. Am. Stat. Assoc.*, **59**: 863-866.
- , 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, U. S. Department of Commerce, Springfield, Virginia.
- , 1981, Constrained Expected Likelihood Estimates of Precision Using Grubbs Technique for Two Measurement Methods, *Nucl. Mater. Manage.*, **X 2**: 34-39.
- , 1983, Special Problems in Measurement Variance Estimation and Hypothesis Testing: The Measurement of the Same n Items by Each of N Methods, *Mathematical and Statistical Methods in Nuclear Safeguards*, Harwood Academic Publishers, London, pp. 57-94.
- Krick, M. S., 1980, *Neutron Multiplication Corrections for Passive Thermal Neutron Well Counters*, LA-8460-MS, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.
- Krutchkoff, R. G., 1967, Classical and Inverse Regression Methods of Calibration, *Technometrics*, **9**: 425-439.
- Ku, H. H. (Ed.), 1969, *Precision Measurement and Calibration*, NBS Special Publication 300, Vol. 1, U. S. Government Printing Office, Washington, D.C.
- Maloney, C. J., and S. C. Rastagi, 1971, Significance Test for Grubbs Estimators, *Biometrics*, **26**: 671-676.
- Mandel, J., 1957, Fitting a Straight Line to Certain Types of Cumulative Data, *J. Am. Stat. Assoc.*, **52**: 552-566.
- Mullen, M. F., 1982, *Selected Topics in Calibration of Nuclear Materials Accountability Tanks*, STR-132, Dept. of Safeguards, International Atomic Energy Agency, Vienna.
- New Brunswick Laboratory, 1983, *Safeguards Analytical Laboratory Evaluation Program (SALE): 1981 Annual Report. Appendix II. Statistical Evaluation Procedures for Bimonthly Data*, NBL-305, New Brunswick Laboratory, Argonne, Illinois.
- Stewart, K. B., 1975, Some Statistical Aspects of Bias Corrections, *Nucl. Mater. Manage.*, Vol. IV, **2**: 20-25.
- , 1980, GROUT: A Computer Program to Test for Outliers and to Estimate Measurement Variance Components (unpublished report), Pacific Northwest Laboratories, Richland, Washington.

CHAPTER 16

Evaluation of Accounting Results

16.0 INTRODUCTION

As discussed in Chapter 13, mass balance accounting is based on a periodic comparison of the book inventory to the physical inventory. At the end of each material balance period an estimate of the amount of material present in the material balance area is subtracted from an estimate of the amount that should be present to obtain an estimate of the inventory difference. The cumulative inventory difference is the sum of the inventory differences for a number of consecutive periods and can be considered as an inventory difference over an extended period. The estimated values of the inventory difference and cumulative inventory difference and the associated estimates of their variances and covariances are the basic input to the evaluation of material accounting results. They provide the basis for both a quantitative assessment of the state of material control and the estimation and testing of the parameters of various loss models.

16.1 AN OVERVIEW OF THE EVALUATION PROBLEM

16.1.1 Scope of the Problem

In one sense the evaluation process is very simple. The estimated inventory difference from the material accounting process is compared to an appropriate multiple of its standard error of estimate. If this critical value is exceeded, the hypothesis of a true inventory difference of zero is rejected and the observed inventory difference is declared "significant." Steps are then taken to investigate the causes of the presumed loss. On the other hand, it has been repeatedly pointed out that the estimated inventory difference is a complex combination of many individual results, each subject to a variety of errors. Reported inventory differences can arise from many causes, such as material that is present but not included in the physical inventory, recording errors, bookkeeping errors, and gross mistakes in measurement, as well as the usual errors of measurement. Such contributions to the variability in an inventory difference estimate are not usually included in an estimate of the variance due to measurement error, and their presence often confuses the interpretation of estimated inventory differences. The initial motivation for material balance accounting was to maintain control over the accounting for nuclear materials by detecting

and eliminating nonmeasurement contributions. Error variance estimates reflecting only measurement error are appropriate to determine the need for investigating accounting discrepancies. Problems of interpretation arise only when inventory differences and other accounting indices are being used to make decisions concerning deliberate loss or diversion. It is necessary in this case to include in the variance estimate any contribution from nonmeasurement sources not eliminated from the accounting process. In particular, estimates of losses that are known but not directly measurable—such as process holdup or accumulation on filters—can be considered as another “measurement” whose variability should be included as an additional source of error.

A somewhat different problem is the estimation or testing of the total effect of accounting errors from a sample. This situation arises when compliance with licensing requirements or international agreements is being verified. Statistical techniques for estimating the effect of recording errors and gross mistakes in measurement under such circumstances have had only limited attention both in general accounting practice and in material accounting. The distribution of such errors tends to be nonnormal, so that appropriate contributions to the variance of an estimate are difficult to estimate. Control practices usually use attribute sampling procedures of the type considered in Chapter 17. While the error estimation techniques considered in Chapter 15 are seldom directly applicable, few alternatives are available. The problem of reporting and recording mistakes in material accounting has been examined by Jaech and Hough (1983), and some techniques for estimating the contribution to the verification process to be considered in Chapter 17 have been suggested. Methods of estimating the overstatement in an accounting population from an audit sample in the absence of measurement error have been reported (Fienberg, Neter, and Leitch, 1977; Neter and Loebhecke, 1977). For further discussion, see Section 17.2.

As discussed in Section 14.3.4, estimates of the measurement error variances—especially those associated with bias effects or bias correction effects—may be based on very few equivalent degrees of freedom and hence, subject to large variability. This also can seriously affect the validity of statistical testing procedures. The final tests of an observed inventory difference or cumulative inventory difference, while apparently simple and straightforward, are the culmination of a complex estimation process involving many assumptions. While experience has shown the usefulness of proceeding with the evaluation techniques described in the following sections, the user should be aware of the complexities involved and the approximate nature of many of the procedures.

The remainder of this section defines the input data and loss models in the form they have generally been used in developing the extensive literature on the evaluation of inventory differences. The next section considers

the application of the statistical hypothesis testing procedures of Chapter 4 to the routine testing and monitoring of inventory differences and cumulative inventory differences. The final section considers specialized estimation procedures useful when sequences of inventory differences are available. These are particularly applicable when unit processes or other well-defined process areas are being monitored on a frequent or near-real-time basis. Examples are daily inventory taking and inventory difference calculations at the dissolver area in a reprocessing plant or at a rod loading station in a MOX plant and weekly inventory taking and inventory difference calculations across the process area of a reprocessing plant.

16.1.2 Defining the Components of an Inventory Difference

The inventory difference of a single material balance period was defined in Equation 13.1 as

$$ID = BI + A - R - EI \quad (16.1)$$

where the terms on the right are the beginning inventory, the additions to the inventory, the removals from the inventory, and the ending inventory, respectively. The estimated inventory difference is a linear combination of observations on each of these four components, and hence, a linear combination of accounting results.

As pointed out in Chapter 14, estimates of individual components as defined in Equation 16.1 are not necessarily independent. There are several ways in which these components can be redefined to emphasize certain dependencies and contrasts (in the sense of Section 5.1.4.2). The expression of the inventory difference in the form

$$ID = (BI + A) - (R + EI) \quad (16.2)$$

in which all positive terms in the ID equation are in the first component and all negative terms are in the second, indicates the magnitudes of the two quantities being differenced to obtain the inventory difference. When treating consecutive inventory differences as sequential observations, it is sometimes convenient to consider the net transfer $T = A - R$ as the observed change in two successive inventories, in which case

$$ID = BI + T - EI \quad (16.3)$$

This is the basic form used for estimation and testing procedures in which both the “book” inventory $BI + T$ and the “physical” inventory EI are considered as estimates of the true inventory at the end of the period—i.e., T is the “change of state” of the inventory over the material balance period. When considering a sequence of n inventory differences, it is convenient to replace the symbols BI and EI by the single set of $n + 1$ inventory symbols I_i , $i = 0, 1, \dots, n$, which represents the inventory at the end of the i^{th} period. Without this change the symbols EI_i and BI_{i+1} would represent the same quantity. In this notation the inventory difference for the i^{th} period in a sequence of n is given by

$$ID_i = I_{i-1} + A_i - R_i - I_i = I_{i-1} + T_i - I_i \quad (16.4)$$

The actual notation used in the literature on evaluation of inventory differences to designate these components varies widely.

The individual components required to form a sequence of n inventory differences using Equation 16.4 are the $2n + 1$ quantities I_i , $i = 0, 1, \dots, n$, and $T_i = A_i - R_i$, $i = 1, \dots, n$. Estimation and testing procedures for inventory differences require that the variances and covariances between the estimates of all these quantities be modeled and estimated. As noted in Chapter 14, a covariance between two of these estimates can occur for two reasons:

1. The identical measurement result may occur in two components, resulting in an exact replication or an exact cancellation of all error effects in the measurement. For example, material received during a material balance period may be contained in the additions and in the ending inventory at identical measured values which cancel exactly (Equation 16.2). This will induce a positive covariance between estimates of the additions and ending inventory as contained in accounting reports. This dependence can be eliminated from a given inventory difference by redefining the components to eliminate these unchanged results.

2. The same error effect may apply to one or more measurements in each of the two components, resulting in the exact replication of one contribution to the total error effect. This is the case, for example, when a scale with a fixed calibration error is used to measure all or part of both shipments and ending inventory. Methods for determining or eliminating the reduction in variability in a linear combination of strata components due to common error effects were considered in Section 14.3.5.

When considering the second source of covariance between components, it is important to recognize that there is not necessarily a relationship between a material balance period and the time period for which the effects β_r' , defined in Section 14.2.3, are constant. This was discussed

briefly in Section 14.3.1 with respect to the definition of error effects arising from the weighing operation in Example 14.1. An important distinction is between the use of a material balance period as an arbitrary period for defining a unique random error effect and the situation where the random error effect is associated with some well-defined time period or set of measurements other than a material balance period. It is only in the latter case that a covariance will exist between two successive net transfers or inventories because the same error effect is part of the total measurement error for results in both components. Effects which are assumed to be randomly chosen for each material balance period, similar to the interclass effects in an analysis of variance, do not contribute to the covariance between components for different periods even though they are constant for all measurements within the period.

Example 16.1 For a certain low-enriched fuel fabrication facility, inventories are taken and material balances computed every 6 months. Assume that the expected (true) values of the flows and inventories for the year 1983 in metric tons of uranium were as follows:

TABLE 16.1
Uranium Flows and Inventories in 1983

| Component | | Amount
(metric tons) |
|-------------------------|-----------|-------------------------|
| Inventory, 12/31/82 | (I_0) | 70 |
| Inventory, 6/30/83 | (I_1) | 40 |
| Inventory, 12/31/83 | (I_2) | 16 |
| Additions, 1/1-6/30/83 | (A_1) | 56 |
| Additions, 7/1-12/31/83 | (A_2) | 44 |
| Removals, 1/1-6/30/83 | (R_1) | 86 |
| Removals, 7/1-12/31/83 | (R_2) | 68 |

The true inventory differences are calculated from Equation 16.1 or 16.4 to be zero. In practice, as will be seen in Example 16.4, the presence of measurement errors in the actual accounting results used to determine the component totals will lead to an observed nonzero inventory difference even in the absence of loss or deliberate diversion.

A campaign involving 90 tons of material was in process as of December 31, 1982. Another similar campaign involving 80 tons of material was initiated during the first half of the year and completed during

the second half. Specifically, of the 70 tons of material from the first campaign on hand as of December 31, 1982, 30 were in the form of finished fuel components, 12 were in process, and 28 tons were in the form of feed material not yet introduced into the process. The receipts during the first 6 months consisted of 20 additional tons of feed to complete the first campaign and 36 tons of feed for the second campaign. All of the latter was still in the June 30, 1983, inventory in some form, along with 4 tons of material in unrecovered scrap from the previous campaign. In particular, the 18 tons of material received but not yet introduced into the process were included in this inventory at the same value they were included in the first period receipts. The shipments during the first 6 months consisted of the 86 tons of material in finished fuel elements from the first campaign, 30 of which were shipped at exactly the same measured value as they were included in the inventory at the beginning of the year. During the last half of the year the remaining 44 tons of material for the second campaign were received and processed. The 6 tons of finished product material in the inventory on June 30, 1983, plus an additional 56 tons of material in fuel elements produced during the second half of the year were shipped along with 6 tons of material from recovered scrap. The inventory at the end of the year consisted of 14 tons of material in finished fuel elements and 2 tons of material in unrecovered scrap.

Each of the inventories is composed of one or more of the following material types:

1. Feed material not yet introduced into the process which is in the inventory at the same value used to determine receipts during the previous period.
2. Material in process.
3. Unrecovered scrap.
4. Material in fuel components (rods or final assemblies) included in inventory at exactly the established content at which it will be included in the shipments for some following period.

The breakdown of each inventory into these four strata is as follows:

| Inventory | Stratum | | | |
|-----------|---------|----|---|----|
| | 1 | 2 | 3 | 4 |
| 12/31/82 | 28 | 12 | — | 30 |
| 6/30/83 | 18 | 12 | 4 | 6 |
| 12/31/83 | — | — | 2 | 14 |

These four inventory strata, the single stratum for receipts, and the two strata for shipments (product and recovered scrap) involve five types of

material whose measurement errors are assumed to arise from the following simplified error structure (Section 14.1.3):

1. Measurements of the receipts of feed material contain individual error effects with a relative standard deviation of 0.5% associated with each 2 tons of uranium per cylinder received and a constant error effect with a relative standard deviation of 0.1% associated with the total receipts.

2. Product shipment measurements contain individual error effects with a relative error effect of 0.25% associated with each 400 kg of uranium in a finished fuel element and a constant error effect with a relative standard deviation of 0.05% associated with the total product shipment.

3. Measurements of the shipment of recovered scrap contain individual error effects with a relative standard deviation of 0.7% associated with each 500 kg shipment and a constant error effect with a relative standard deviation of 0.2% associated with the total shipment.

4. Estimates of the total material in the process inventory contain errors with a relative standard deviation of 0.3%.

5. Measurements of the material content of unrecovered scrap contain individual error effects with a relative standard deviation of 3% associated with each scrap batch containing a nominal 100 kg of uranium, and a constant error effect of 0.8% associated with the total stored unrecovered scrap.

These assumptions are summarized in Table 16.2 using the notation of Section 14.3.1.

TABLE 16.2
Assumed Error Parameters

| Material type | Relative standard deviation (%) | |
|----------------------|---------------------------------|----------|
| | θ | δ |
| Feed material | 0.1 | 0.5 |
| Finished elements | 0.05 | 0.25 |
| Scrap shipments | 0.2 | 0.7 |
| In-process inventory | — | 0.3 |
| Unrecovered scrap | 0.8 | 3.0 |

From this description of the process flows and inventories and the behavior of the measurement errors, the standard errors of estimate for the components given in Table 16.1 can be computed using the methods developed in Chapter 14. For example, because there are three material types involved, two with two error sources and one with a single source, from Equation 14.47 the variance of an observed value of I_0 would be

$$\begin{aligned}\text{Var}(\hat{I}_0) &= (28,000)^2 (0.001)^2 + 14(2000)^2 (0.005)^2 \\ &\quad + (12,000)^2 (0.003)^2 \\ &\quad + (30,000)^2 (0.005)^2 + 75(400)^2 (0.0025)^2 \\ &= 3780 \text{ (kg)}^2\end{aligned}$$

For A_1 there is only one material type, and

$$\begin{aligned}\text{Var}(\hat{A}_1) &= (56,000)^2 (0.001)^2 + 28(2000)^2 (0.005)^2 \\ &= 5936 \text{ (kg)}^2\end{aligned}$$

For R_2 there are two material types, each with two error sources, so that

$$\begin{aligned}\text{Var}(\hat{R}_2) &= (62,000)^2 (0.0005)^2 + 155(400)^2 (0.0025)^2 \\ &\quad + (6000)^2 (0.002)^2 + 12(500)^2 (0.007)^2 \\ &= 1407 \text{ (kg)}^2\end{aligned}$$

The variances and standard errors for all the components in Table 16.1 are given in Table 16.3.

TABLE 16.3
Standard Errors of Flows and Inventories

| Component | Variance (kg) ² | Standard error (kg) |
|-----------|----------------------------|---------------------|
| I_0 | 3780 | 61.5 |
| I_1 | 3928 | 62.7 |
| I_2 | 520 | 22.8 |
| A_1 | 5936 | 77.0 |
| A_2 | 4136 | 64.3 |
| R_1 | 2064 | 45.4 |
| R_2 | 1407 | 37.5 |

In this example, estimates of I_0 and R_1 will be dependent and will involve a covariance because of the 30 tons of finished fuel elements in the December 31, 1982, inventory. These elements were shipped during the first 6 months and occur in both estimates at the same measured value. The covariance is (Section 14.3.5)

$$\begin{aligned}\text{Cov}_1(\hat{I}_0, \hat{R}_1) &= (30,000)^2(0.0005)^2 \\ &+ (75)(400)^2(0.0025)^2 \\ &= 300(\text{kg})^2\end{aligned}$$

The subscript indicates that the covariance is due to the first of the two sources discussed earlier in this section. Estimates of I_1 and A_1 will be dependent because of the 18 tons of receipts still included in the inventory estimate at the same measured value as in the receipts. The estimate of the covariance is

$$\begin{aligned}\text{Cov}_1(\hat{A}_1, \hat{I}_1) &= (18,000)^2(0.001)^2 \\ &+ (9)(2000)^2(0.005)^2 \\ &= 1224(\text{kg})^2\end{aligned}$$

Because there was no feed material in inventory on December 31, 1983, \hat{A}_2 and \hat{I}_2 contain no dependency of this type in this case. On the other hand, the June 30, 1983, inventory and the shipments during this second 6 months have in common 6 tons of finished fuel elements, so that

$$\begin{aligned}\text{Cov}_1(\hat{I}_1, \hat{R}_2) &= (6000)^2(0.0005)^2 \\ &+ (15)(400)^2(0.0025)^2 \\ &= 24(\text{kg})^2\end{aligned}$$

Assuming that the measurement of in-process material is replicated at each inventory, and that there is no common feed material, product, or unrecovered scrap in the three inventories, there is no further dependency due to common measurement results in this example. This is not generally true; in fact, inventory dependencies due to the inclusion of material in several successive inventories at identical values are quite common. In particular, this example assumes that the 2 tons of unrecovered scrap in the December 31, 1983, inventory is completely different from the 4 tons in the June 30, 1983, inventory—i.e., that all the unrecovered scrap in the June 30, 1983, inventory was part of the 6 tons recovered and shipped, so that none remains on inventory at an identical value.

The above dependencies arise because the same material is included in two components. Dependencies also arise because different material of the same material type occurs in two components system with a common constant error effect associated with the measurement system. For purposes of illus-

tration, assume that for each material type a single constant error effect applied to all measurements during the year. Five components in Table 16.1 involve feed material; thus, there are 10 possible covariances due to feed material, each based on the relative standard deviation of the common error effect given in Table 16.2 and the amounts of different feed material in each component. For example,

$$\begin{aligned}\text{Cov}_2(\hat{I}_0, \hat{A}_1) &= (28,000)(56,000)(0.001)^2 \\ &= 1568 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Cov}_2(\hat{I}_0, \hat{I}_1) &= (28,000)(18,000)(0.001)^2 \\ &= 504 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Cov}_2(\hat{A}_1, \hat{I}_1) &= (38,000)(18,000)(0.001)^2 \\ &= 684 \text{ (kg)}^2\end{aligned}$$

The subscript 2 indicates that the covariance is induced by a common constant error effect. The last of these is in addition to the previously computed covariance $\text{Cov}_1(\hat{A}_1, \hat{I}_1)$ between \hat{A}_1 and \hat{I}_1 associated with the 18 tons of identical feed material in the two components. Similarly, product material is contained in five components and unrecovered scrap in two. The errors in the in-process inventory estimates are uncorrelated.

From the definition of ID_1 in Equation 16.4 and using Equation 2.123,

$$\begin{aligned}\text{Var}(\hat{ID}_1) &= \text{Var}(\hat{I}_0) + \text{Var}(\hat{A}_1) + \text{Var}(\hat{R}_1) + \text{Var}(\hat{I}_1) \\ &\quad + 2 \text{Cov}(\hat{I}_0, \hat{A}_1) - 2 \text{Cov}(\hat{I}_0, \hat{R}_1) - 2 \text{Cov}(\hat{I}_0, \hat{I}_1) \\ &\quad - 2 \text{Cov}(\hat{A}_1, \hat{R}_1) - 2 \text{Cov}(\hat{A}_1, \hat{I}_1) + 2 \text{Cov}(\hat{R}_1, \hat{I}_1)\end{aligned}$$

In general, the covariance between two components is the sum of the covariance induced by identical measurements and the covariance induced by common constant error effects. For example,

$$\text{Cov}(\hat{A}_1, \hat{I}_1) = \text{Cov}_1(\hat{A}_1, \hat{I}_1) + \text{Cov}_2(\hat{A}_1, \hat{I}_1)$$

To evaluate $\text{Var}(\hat{ID}_1)$, the following additional covariance components based on nonidentical product material measured with common error effects are needed:

$$\begin{aligned}\text{Cov}_2(\hat{I}_0, \hat{R}_1) &= (30,000)(56,000)(0.0005)^2 \\ &= 420(\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Cov}_2(\hat{I}_0, \hat{I}_1) &= (30,000)(6,000)(0.0005)^2 \\ &= 45(\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Cov}_2(\hat{R}_1, \hat{I}_1) &= (86,000)(6,000)(0.0005)^2 \\ &= 129(\text{kg})^2\end{aligned}$$

Substituting from previous computations and Table 16.3,

$$\begin{aligned}\text{Var}(\hat{I}D_1) &= 3780 + 5936 + 2064 + 3928 \\ &\quad + 3136 - (600 + 840) - (1008 + 90) \\ &\quad - 0 - (2448 + 1368) + 258 \\ &= 12,748 (\text{kg})^2\end{aligned}$$

so that the standard error of estimate is 112.9 kg.

For purposes of inventory difference evaluation, the dependence arising from the use of identical measured values in both inventory and flow components can be eliminated by redefining the components of the inventory difference so that feed material in an ending inventory is considered as an addition during the following period, and product material in a beginning inventory as a removal during the previous period. Specifically, let

$$A_i = A_{1i} + A_{2i}$$

where

A_{1i} = material received and processed in period i

A_{2i} = material received during period i but not yet processed

and similarly

$$R_i = R_{1i} + R_{2i}$$

where

- R_{1i} = material produced (including scrap) and removed in period i
 R_{2i} = material produced (including scrap) in period $i-1$ but not yet removed

Then

$$I_i = A_{2i} + R_{2,i+1} + I'_i$$

where I'_i is the material in process plus the unrecovered scrap. From Equation 16.4,

$$\begin{aligned} ID_i &= I_{i-1} + A_i - R_i - I_i \\ &= A_{2,i-1} + R_{2i} + I'_{i-1} + A_{1i} + A_{2i} \\ &\quad - R_{1i} - R_{2i} - A_{2i} - R_{2,i+1} - I'_i \end{aligned}$$

Define

$$\begin{aligned} A'_i &= A_{2,i-1} + A_{1i} \\ R'_i &= R_{1i} + R_{2,i+1} \end{aligned}$$

Then it follows that

$$ID_i = I'_{i-1} + A'_i - R'_i - I'_i$$

These new material balance components are defined in terms of the process rather than the actual presence of the material at the facility. The advantage is that the duplication of the quantities A_{2i} and R_{2i} in the original components has been eliminated and the new components are independent linear combinations of the five basic quantities, A_{1i} , A_{2i} , R_{1i} , R_{2i} , and I'_i . Making the boundaries of the material balance area reflect process changes rather than geographical or administrative areas has tended to associate the components of the material balance with key measurements rather than changes in location or responsibility.

In some cases, material received during period i might not be processed for several periods, and material produced might not be shipped for several periods. Both of these circumstances lead to inventory which is unchanged for one or more periods. The affected inventories must be redefined by

removing any such inactive material from both affected inventories. Such changes due to a common inventory are in addition to the quantities A_{21} and R_{21} and would be subtracted from the redefined inventories I_1' .

Example 16.2 Continuing with Example 16.1, the basic independent quantities defining the material balance components for the year are:

- I_0' = 12 tons of uranium in material in process as of December 31, 1982
- I_1' = 16 tons of uranium in material and unrecovered scrap in process as of June 30, 1983
- I_2' = 2 tons of uranium in unrecovered scrap in process as of December 31, 1983
- A_{20} = 28 tons of uranium in feed material on hand as of December 31, 1982, but not processed
- A_{11} = 38 tons of uranium in feed material received and processed in the first 6 months
- A_{21} = 18 tons of uranium in feed material received in the first 6 months but not processed
- A_{12} = 44 tons of uranium in feed material received and processed in the second 6 months
- A_{22} = 0 tons of uranium in feed material received in the second 6 months but not processed
- R_{21} = 30 tons of uranium in fuel produced but not shipped as of December 31, 1982
- R_{11} = 56 tons of uranium in fuel produced and shipped in the first 6 months
- R_{22} = 6 tons of uranium in fuel produced in the first 6 months but shipped in the second 6 months
- R_{12} = 56 tons of uranium in fuel produced and shipped in the second 6 months, plus 6 tons of uranium recovered from scrap and shipped during the second 6 months
- R_{23} = 14 tons of uranium in fuel produced in the second 6 months but shipped in the next period

The new components are given in Table 16.4. Note that R_{21} , the 30 tons of material produced but not shipped at the beginning of the year, is not involved in these components. If there had been feed material in storage as of the end of the year (A_{22}), it would also have been excluded from the process material balance.

TABLE 16.4
Redefined Material Balance Components

| Component | Amount (tons) |
|--|------------------|
| In-process inventory, 12/31/82 (I ₀) | 12 |
| In-process inventory, 6/30/83 (I ₁) | 16 (12 + 4) |
| In-process inventory, 12/31/83 (I ₂) | 2 |
| Additions, 1/1–6/30/83 (A ₁) | 66 (28 + 38) |
| Additions, 7/1–12/31/83 (A ₂) | 62 (18 + 44) |
| Removals, 1/1–6/30/83 (R ₁) | 62 (56 + 6) |
| Removals, 7/1–12/31/83 (R ₂) | 76 (56 + 6 + 14) |

The components as defined in Table 16.4 do not contain identical measured values for common items, so that this source of covariance has been eliminated. Because of the magnitude of the throughput and the time intervals involved, there is a high turnover rate, so that the covariances between the successive in-process inventories and between these inventories and the additions to and removals from the process are frequently negligible in practice. However, because fixed effects common to both material balance periods may constitute a significant part of the error present in additions and removals, the consequent covariance between the estimated amounts in the various strata constituting the receipts and removals may be appreciable. Consequently, the methods of Section 14.3.5 must be used in determining the standard error estimates for the combinations of comparable strata in these components (Example 14.5).

From Table 16.2 and Equation 14.47, the variances of the estimated redefined components are as follows:

$$\begin{aligned}\text{Var}(\hat{I}_0) &= (12,000)^2(0.003)^2 \\ &= 1296 (\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{I}_1) &= (12,000)^2(0.003)^2 + (4000)^2(0.008)^2 \\ &\quad + 40(100)^2(0.03)^2 \\ &= 2680 (\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{I}_2) &= (2000)^2(0.008)^2 + 20(100)^2(0.03)^2 \\ &= 436 (\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{A}'_1) &= (66,000)^2(0.001)^2 + 33(2000)^2(0.005)^2 \\ &= 7656 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{A}'_2) &= (62,000)^2(0.001)^2 + 31(2000)^2(0.005)^2 \\ &= 6944 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{R}'_1) &= (62,000)^2(0.0005)^2 + 155(400)^2(0.0025)^2 \\ &= 1116 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{R}'_2) &= (70,000)^2(0.0005)^2 + (175)(400)^2(0.0025)^2 \\ &\quad + (6000)^2(0.002)^2 + (12)(500)^2(0.007)^2 \\ &= 1691 \text{ (kg)}^2\end{aligned}$$

\hat{ID}_1 is given by

$$\hat{ID}_1 = \hat{I}'_0 + \hat{A}'_1 - \hat{R}'_1 - \hat{I}'_1$$

There are no dependencies between the redefined elements due to measurement results common to two components and no duplication of measurement methods (material types) in more than one stratum except for the in-process inventory estimates, which are assumed to be independent. Hence,

$$\begin{aligned}\text{Var}(\hat{ID}_1) &= \text{Var}(\hat{I}'_0) + \text{Var}(\hat{A}'_1) \\ &\quad + \text{Var}(\hat{R}'_1) + \text{Var}(\hat{I}'_1) \\ &= 12,748 \text{ (kg)}^2\end{aligned}$$

which is the result obtained in Example 16.1. Similarly,

$$\begin{aligned}\text{Var}(\hat{ID}_2) &= \text{Var}(\hat{I}'_1) + \text{Var}(\hat{A}'_2) \\ &\quad + \text{Var}(\hat{R}'_2) + \text{Var}(\hat{I}'_2) \\ &\quad - 2 \text{Cov}(\hat{I}'_1, \hat{I}'_2)\end{aligned}$$

The single nonzero covariance in this case is due to the assumed presence of a common error effect in the measurement of unrecovered scrap in process, so that

$$\begin{aligned}\text{Cov}(\hat{I}'_1, \hat{I}'_2) &= (4000)(2000)(0.008)^2 \\ &= 512 (\text{kg})^2\end{aligned}$$

and hence,

$$\begin{aligned}\text{Var}(\hat{ID}_2) &= 10,727 (\text{kg})^2 \\ \text{SE}(\hat{ID}_2) &= 103.6 \text{ kg}\end{aligned}$$

In addition to the possible correlations between component estimates, successive inventory differences are dependent because they always, by definition, have one inventory component in common. However, this causes no difficulty if it is remembered that the sum of two successive inventory differences is given by

$$\begin{aligned}ID_i + ID_{i+1} &= (I_{i-1} + T_i - I_i) + (I_i + T_{i+1} - I_{i+1}) \\ &= I_{i-1} + T_i + T_{i+1} - I_{i+1}\end{aligned}\tag{16.5}$$

so that the variance of the cumulative difference involves only the variances and covariances of the four quantities I_{i-1} , T_i , T_{i+1} , and I_{i+1} . Note that the variance of the intermediate inventory has no effect on the variance of the sum of the inventory differences.

Theoretical investigations of procedures for evaluating inventory differences often assume the independence of the estimates of inventories and net transfers. Under this assumption, the only estimates of inventory differences that are not independent are estimates of successive inventory differences, which are dependent because (as noted in the previous paragraph) they have the estimate of one inventory component in common. The covariance of these successive estimates is the negative of the variance of the estimate of the common inventory component. Since experience has shown that the covariances between components due to common error effects tend to be small when effective control practices are followed, these independence assumptions are probably justified for theoretical investigations of estimation procedures. In practice, the magnitudes of the variances and covariances estimated using the formulations of Chapter 14 will vary widely with the nature and amount of material processed or transferred and, as pointed out there, individual treatment of the sources of measurement error is required to determine the specific variances and covariances for each observed inventory difference from available estimates of the variance components from each source.

Example 16.3 Continuing the example of 16.1 and 16.2, the covariance between

$$\hat{ID}_1 = \hat{I}'_0 + \hat{A}'_1 - \hat{R}'_1 - \hat{I}'_1$$

and

$$\hat{ID}_2 = \hat{I}'_1 + \hat{A}'_2 - \hat{R}'_2 - \hat{I}'_2$$

arises from four sources. The first contribution is due to the common inventory estimate \hat{I}'_1 , and is given by

$$\begin{aligned} \text{Cov}(\hat{ID}_1, \hat{ID}_2) &= -\text{Var}(\hat{I}'_1) \\ &= -2680 \text{ (kg)}^2 \end{aligned}$$

The second component arises from the assumption of a common constant error effect in \hat{A}'_1 and \hat{A}'_2 , and is given by

$$\begin{aligned} \text{Cov}(\hat{A}'_1, \hat{A}'_2) &= (66,000)(62,000)(0.001)^2 \\ &= 4092 \text{ (kg)}^2 \end{aligned}$$

The third component arises from the assumption of a common constant error effect in \hat{R}'_1 and \hat{R}'_2 , and is given by

$$\begin{aligned} \text{Cov}(\hat{R}'_1, \hat{R}'_2) &= (62,000)(70,000)(0.0005)^2 \\ &= 1085 \text{ (kg)}^2 \end{aligned}$$

The fourth component is due to the common error effect in the unrecovered scrap measurements in \hat{I}'_1 and \hat{I}'_2 , which was previously evaluated in Example 16.2 as

$$\text{Cov}(\hat{I}'_1, \hat{I}'_2) = 512 \text{ (kg)}^2$$

Then

$$\begin{aligned} \text{Var}(\hat{ID}_1 + \hat{ID}_2) &= \text{Var}(\hat{ID}_1) + \text{Var}(\hat{ID}_2) \\ &\quad - 2 \text{Var}(\hat{I}'_1) + 2 \text{Cov}(\hat{A}'_1, \hat{A}'_2) \end{aligned}$$

$$\begin{aligned}
& + 2 \text{Cov}(\hat{R}'_1, \hat{R}'_2) + 2 \text{Cov}(\hat{I}'_1, \hat{I}'_2) \\
& = 12,748 + 10,727 - 5360 \\
& \quad + 8184 + 2170 + 1024 \\
& = 29,493 (\text{kg})^2
\end{aligned}$$

and the standard error of estimate for the cumulative inventory difference for the year is 171.7 kg.

More directly, from Equation 16.5 it follows that

$$\hat{I}D_1 + \hat{I}D_2 = \hat{I}'_0 + (\hat{A}'_1 + \hat{A}'_2) - (\hat{R}'_1 + \hat{R}'_2) - \hat{I}'_2$$

For the total additions $\hat{A}'_1 + \hat{A}'_2 = 128$ tons, the variance can be calculated directly from the relative standard deviations in Table 16.2 as

$$\begin{aligned}
\text{Var}(\hat{A}'_1 + \hat{A}'_2) & = (128,000)^2(0.001)^2 \\
& \quad + 64(2000)^2(0.005)^2 \\
& = 22,784 (\text{kg})^2
\end{aligned}$$

The total removal is $\hat{R}'_1 + \hat{R}'_2 = 138$ tons, 132 tons of finished fuel, and 6 tons of recovered scrap. Thus,

$$\begin{aligned}
\text{Var}(\hat{R}'_1 + \hat{R}'_2) & = (132,000)^2(0.0005)^2 \\
& \quad + (330)(400)^2(0.0025)^2 \\
& \quad + (6000)^2(0.002)^2 \\
& \quad + 12(500)^2(0.007)^2 \\
& = 4977 (\text{kg})^2
\end{aligned}$$

Because the estimates \hat{I}'_0 and \hat{I}'_2 are independent of each other and of the estimates of total additions and total removals,

$$\begin{aligned}
\text{Var}(\hat{I}D_1 + \hat{I}D_2) & = \text{Var}(\hat{I}'_0) + \text{Var}(\hat{A}'_1 + \hat{A}'_2) \\
& \quad + \text{Var}(\hat{R}'_1 + \hat{R}'_2) + \text{Var}(\hat{I}'_2)
\end{aligned}$$

$$\begin{aligned} &= 1296 + 22,784 \\ &\quad + 4977 + 436 \\ &= 29,493 \text{ (kg)}^2 \end{aligned}$$

as obtained previously.

16.1.3 Loss Models

In Chapter 14 it was emphasized that an estimate of an inventory difference is the sum, with appropriate signs, of a large number of individual estimates of the content of individual items. These individual estimates are accumulated into estimates of stratum totals, component totals, and inventory differences. The simplest model for an estimated inventory difference is

$$\hat{ID} = L + \epsilon \quad (16.6)$$

where L is the expected value of the inventory difference and ϵ the random error present in the observed result. As discussed in Chapter 13, in the absence of unknown losses, accounting mistakes, and measurement error, L and ϵ should both be identically equal to zero. In a completely controlled situation, where only measurement error effects with zero expected values are present, the expected value L will be identically zero and ϵ will be a random variable with $E(\epsilon) = 0$ and $\text{Var}(\epsilon) = \text{Var}(\hat{ID})$. The basis for this assumption is discussed in Section 14.3.1 in conjunction with the assumptions made in defining procedures for estimating $\text{Var}(\hat{ID})$.

With respect to accounting mistakes, such as recording errors, missed inventory items, or unmeasured wastes, two approaches are possible. If the purpose of the inventory difference evaluation is to maintain a state of control over the accounting process, then this type of contribution to \hat{ID} is assumed to be included in L , and specific testing procedures are designed to discover and eliminate such assignable causes of a nonzero value of L . When the purpose of the evaluation is to discover unknown losses or deliberate diversions of material, it is necessary to assume *a priori* that the accounting process is in control, i.e., that all mistakes and other non-measurement sources of variability have been eliminated or their effects included in the determination of the variance of the random variable ϵ . The control procedures can include identification of and correction for consistently positive (or even negative) losses which are an inherent part of the accounting process. However, as is the case of measurement control

procedures for individual sources of measurement error, it is always preferable to eliminate the need for such corrections whenever possible.

When a sequence of inventory differences is considered, the assumed model is

$$\hat{ID}_i = L_i + \epsilon_i \quad (16.7)$$

where the L_i are parameters and the ϵ_i are random variables with $E(\epsilon_i) = 0$. The ϵ_i will be assumed to have a known variance-covariance matrix. Some theoretical studies have been carried out under the simplifying assumption that the covariance between all estimates of inventories and net transfers is zero, so that $\text{Cov}(\epsilon_i, \epsilon_{i-k}) = 0$ for $k > 1$. Recommended procedures for sequential testing do not make this assumption and are based on a redefined sequence of independent quantities computed using a known variance-covariance matrix (Goldman, 1985).

In connection with the development of procedures for the testing of sequences of inventory difference estimates, a variety of assumptions has been made with respect to the expected values L_i . The nature of the assumptions differs depending on whether the purpose is control of unknown losses or the detection of deliberate diversion. In the first case a common assumption is that the L_i are observed values of a random variable with expected value L and variance σ_L^2 , where L may or may not be assumed to be 0. This model is similar to the random effects model for the between-class effects in a one-way analysis of variance. The special case $\sigma_L^2 = 0$, for which $L_i = L$ for all i , is frequently called the constant loss model. When dealing with deliberate diversion rather than chance losses, the values of the L_i can presumably be chosen by the diverter to maximize his chances of nondetection. This allows a wide variety of diversion patterns, usually subject only to the constraint that the total amount $M = \sum L_i$ diverted over some fixed period equals or exceeds some specified goal. In this case the constant loss model in which an amount $L = M/n$ is diverted each period is sometimes referred to as "protracted" diversion, as opposed to the diversion of an amount M in a single material balance period, which is called "abrupt" diversion. Many specialized tests have been developed to deal with specific diverter strategies as alternatives to the null hypothesis of no diversion. Some of these are discussed in Section 16.3.

16.2 TESTING INVENTORY DIFFERENCES AND CUMULATIVE INVENTORY DIFFERENCES

When the observed inventory difference is considered as a normally distributed random variable with expected value L and known variance, the hypothesis tests of Section 4.2.1 are applicable. Considering the large

number and nature of the individual results involved, the assumption of normality based on the Central Limit Theorem is almost certainly justified. Although there is no practical alternative, the assumption of known variance should be examined more carefully. If the variance of the inventory difference is dominated by several variance components whose estimates are based on small numbers of degrees of freedom (Section 3.5.1), this assumption may be questionable and result in an increased number of unwarranted rejections of the null hypothesis (false positives), because the most common consequence of variance estimation based on small numbers of degrees of freedom is an underestimation of the true magnitude of the variance.

The basic testing procedures for inventory differences and cumulative inventory differences are considered in Section 16.2.1. Section 16.2.2 reviews some applications of decision theory and game theory to the determination of optimum testing procedures.

16.2.1 Hypothesis Testing

When the inventory difference is being used as a control on the accounting process, we are concerned with the composite alternative hypothesis $L \neq 0$ as opposed to the null hypothesis $L = 0$. The control chart procedures of Section 4.7 are directly applicable. The situation has some aspects of both theoretical and empirical control charts. In the sense that their primary use is to promote guidance in establishing and maintaining a state of control, they are empirical; but in the sense that each observed inventory difference has a unique known variance which is used to determine specific control lines, they are theoretical.

In order to establish a state of control such that $L = 0$, the center line of the control chart is chosen to be zero. As noted in the discussion of routine monitoring in Section 15.2.1.1, it has become standard practice in nuclear regulatory applications, as mandated by the Nuclear Regulatory Commission, to use approximate 5% and 1% limits to make control judgments rather than the customary 2σ and 3σ limits of quality control practice.† The justification for this is based on the desire to emphasize investigation of possible assignable causes at the expense of false positives. Further discussion of the criteria for determining action limits and critical values is contained in the next section and in Chapter 12.

The estimates $\hat{SE}(\hat{ID}) = \sqrt{\hat{Var}(\hat{ID})}$ needed to compute these limits are generally determined in one of two ways:

1. By estimating $Var(\hat{ID})$ using the methods of Chapter 14.
2. By multiplying some historical standard deviation relative to the throughput, considered to be representative of good accounting practice for

† In particular, the term "limits of error" has been commonly used for estimates of the 2σ , or approximately 5%, limits on both individual measurements and accounting results.

the process being monitored, by the throughput associated with an estimate $\hat{I}D$. It follows from Equation 16.2 that if this known standard deviation relative to the throughput is defined as δ , and if the throughput T is defined as either $BI + A$ or $R + EI$, then an estimate of the variance is

$$\text{Vâr}(\hat{I}D) \cong 2(\delta T)^2$$

and

$$\hat{S}E(\hat{I}D) \cong \sqrt{2} \delta T$$

The second approach is more applicable to control procedures designed to demonstrate a required or desired level of performance, while the first is more applicable to empirical processes for achieving control (Section 4.7).

The procedure of Section 4.2.1.2 is usually used to test the hypothesis of no loss against the alternative of an unknown loss or a deliberate diversion. The test statistic Z , defined by Equation 4.6, is obtained by replacing the mean \bar{X} and its standard error σ/\sqrt{n} by $\hat{I}D$ and $SE(\hat{I}D)$, respectively, and μ by L . Because the true standard error $SE(\hat{I}D)$ is not known, the procedure is modified by replacing it with an estimate obtained using the approximate procedures of Chapter 14. As noted in Section 14.3.4, information concerning the quality of the standard error estimates obtained in practice is generally not available, and there is little else to do but accept them as known. It should be recognized that the rejection rate under the null hypothesis (false alarm rate) may increase and the probability of detection may decrease when the true standard error is replaced by an estimate. If information is available concerning an appropriate degrees of freedom to assign to the estimate of the standard error, then the parallel procedures of Section 4.2.2 based on Student's t -distribution should be used. This is particularly important if some dominant component of the estimated variance is known to be based on relatively few degrees of freedom.

In this procedure, the hypothesis

$$H_0: L = 0$$

is tested against the one-sided alternative that the loss is some specific positive quantity L_A , so that

$$H_A: L = L_A > 0$$

The modified test statistic is

$$Z = \frac{\hat{ID}}{SE(\hat{ID})} \quad (16.8)$$

For a selected false alarm rate α , the rejection region is $Z \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is determined by

$$\phi(Z < z_{1-\alpha}) = 1 - \alpha$$

and $\phi(z)$ is the cumulative unit normal distribution defined by Equation 2.35 and tabulated in Table A3. From Equation 4.14 the probability β of not rejecting the null hypothesis when the alternative is true is

$$\beta = \phi\left(z_{1-\alpha} - \frac{L_A}{SE(\hat{ID})}\right) \quad (16.9)$$

For a given value of L_A , an estimate of β can be obtained by replacing the known standard error by its estimate as in the testing procedure.

The probability $1 - \beta$ of rejecting the null hypothesis when the alternative hypothesis L_A is true (defined as the power of the test in Section 4.2.1.3) is usually referred to in nuclear material accounting applications as the "probability of detection" of a loss of size L_A .

Because the cumulative inventory difference is simply the inventory difference for an extended period, the identical hypothesis testing procedure is applicable. The hypothesis

$$H_0: M = 0$$

is tested against the alternative

$$H_A: M = M_A > 0$$

using the test statistic

$$Z = \frac{\hat{CID}}{SE(\hat{CID})} \quad (16.10)$$

and the rejection region $Z > z_{1-\alpha}$, where Z is assumed to be $N(0,1)$. The

power to detect an alternative M_A is given by $1 - \beta$, where

$$\beta = \phi \left(z_{1-\alpha} - \frac{M_A}{SE(\hat{C}\hat{I}D)} \right) \quad (16.11)$$

This can again be estimated by replacing $SE(\hat{C}\hat{I}D)$ by $\hat{SE}(\hat{C}\hat{I}D)$.

Example 16.4 Assume that for the material balance periods defined in Example 16.1 the actual reported results were:

TABLE 16.5
Reported Results

| Stratum | Estimated total content
(kg uranium) |
|------------------------------------|---|
| Process material inventory | |
| 12/31/82 | 11,956 |
| 6/30/83 | 12,087 |
| 12/31/83 | — |
| Unrecovered scrap inventory | |
| 12/31/82 | — |
| 6/30/83 | 4,030 |
| 12/31/83 | 1,687 |
| Feed material | |
| In inventory, 12/31/82 | 27,943 |
| Received and processed, 1/1–6/30 | 37,922 |
| In inventory, 6/30/83 | 18,006 |
| Received and processed, 7/1–12/31 | 44,014 |
| In inventory, 12/31/83 | — |
| Finished fuel | |
| In inventory, 12/31/82 | 30,004 |
| Produced and shipped, 1/1–6/30 | 56,015 |
| In inventory, 6/30/83 | 6,002 |
| Produced and shipped, 7/1–12/31 | 56,024 |
| In inventory, 12/31/83 | 14,006 |
| Recovered scrap shipments | |
| 1/1–6/30 | — |
| 7/1–12/31 | 6,011 |

From these results for individual strata the inventories, additions, and

removals on a facility basis as defined in Example 16.1 were computed and are given in Table 16.6.

TABLE 16.6
Reported Components of the Material Balance

| Component | | Estimated amount |
|-----------|-----------|------------------|
| Inventory | 12/31/82 | 69,903 |
| Inventory | 6/30/83 | 40,125 |
| Inventory | 12/31/83 | 15,693 |
| Additions | 1/1-6/30 | 55,928 |
| Additions | 7/1-12/31 | 44,014 |
| Removals | 1/1-6/30 | 86,019 |
| Removals | 7/1-12/31 | 68,037 |

From these component results the estimates of the inventory differences are computed as

$$\hat{ID}_1 = 69,903 + 55,928 - 86,019 - 40,125 = -313$$

$$\hat{ID}_2 = 40,125 + 44,014 - 68,037 - 15,693 = 409$$

The results for the reformulated components based on a process material balance are given in Table 16.7. The standard errors of estimate for these components as computed in Examples 16.2 and 16.3 are also shown for reference.

TABLE 16.7
Estimated Amounts and Estimated Standard Errors for the Reformulated Components of the Material Balance

| Component | | Estimated amount | Standard error of estimate |
|-----------------------|------------|------------------|----------------------------|
| In-process inventory | 12/31/82 | 11,956 | 36 |
| In-process inventory | 6/30/83 | 16,117 | 52 |
| In-process inventory | 12/31/83 | 1,687 | 21 |
| Additions to process | 1/1-6/30 | 65,865 | 87 |
| Additions to process | 6/30-12/31 | 62,020 | 83 |
| Additions to process | 1/1-12/31 | 127,885 | 151 |
| Removals from process | 1/1-6/30 | 62,017 | 33 |
| Removals from process | 7/1-12/31 | 76,041 | 41 |
| Removals from process | 1/1-12/31 | 138,058 | 71 |

The estimates of the inventory differences computed from Table 16.7 are

$$\hat{ID}_1 = 11,956 + 65,865 - 62,017 - 16,117 = -313$$

and

$$\hat{ID}_2 = 16,117 + 62,020 - 76,041 - 1,687 = 409$$

which, as expected, are identical to those computed from Table 16.6. The standard errors of these estimates and their sum were estimated in Examples 16.2 and 16.3 to be:

| Period | \hat{ID} | $\hat{SE}(\hat{ID})$ |
|----------------------|------------|----------------------|
| 1/1-6/30 (ID_1) | -313 | 113 |
| 7/1-12/31 (ID_2) | 409 | 104 |
| 1/1-12/31 (CID) | 96 | 172 |

Note that in Table 16.7 all of the basic material balance components, except the unrecovered scrap in inventory at the end of the year, have standard errors consistent with the 0.1 to 0.3% standard deviation due to measurement errors expected in fuel fabrication plants processing low enriched uranium. However, all the inventory difference estimates are differences of much larger amounts, and such estimates of the relative standard deviation are not directly meaningful with respect to the observed material balances. While the net transfers used in Equation 16.3 are not, like the inventory difference, expected to be zero, they will also usually be small compared to the combination of beginning inventory plus additions, or removals plus ending inventory, with respect to which historical or desirable relative error estimates are defined.

The test statistics obtained by dividing the estimated inventory differences and the cumulative inventory difference by their estimated standard errors are:

$$z_1 = -313/113 = -2.77$$

$$z_2 = 409/104 = 3.93$$

$$z = 96/172 = 0.56$$

Both z_1 and z_2 exceed the critical value $z_{1-0.05} = z_{0.95} = 1.645$. The fact that they are relatively large but opposite in sign suggests some

problem with the intermediate inventory. However, for the process throughputs per period of approximately 78 metric tons (beginning in-process inventory plus additions or ending in-process inventory plus removals) observed here, the estimated standard errors are only about 0.14% of the throughput, which is somewhat smaller than accepted standards for this type of facility. This suggests the alternative additional possibility that bias contributions to the estimated standard error from some sources may have gone unrecognized, leading to underestimation of the standard errors and inflated values of the test statistics.

The procedure for examining the power of these tests is illustrated for the case of $\hat{C}\hat{I}\hat{D}$. The rejection region for the test of $M = 0$ for $\alpha = 0.05$ is $z \geq 1.645$. From Equation 16.11, β is given by

$$\beta = \Phi \left(1.645 - \frac{M_A}{SE(\hat{C}\hat{I}\hat{D})} \right)$$

Using the estimate $\hat{SE}(\hat{C}\hat{I}\hat{D}) = 172$, the probability $1 - \beta$ of rejecting the hypothesis that no loss occurred is plotted as a function of the true loss M_A in Figure 16.1. This power curve describes the loss detection sensitivity of this particular determination of an inventory difference. In particular, the probability of detecting a loss of $M_A = 566$ kg is 0.95 ($\beta = 0.05$). The probability is only 0.5 of detecting a loss of 283 kg.

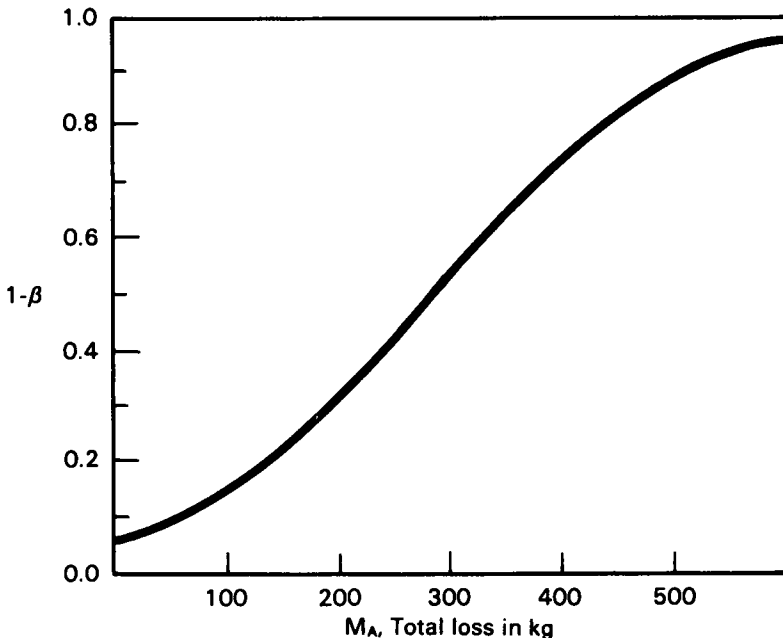


Figure 16.1. Power curve for Example 16.4.

The ability of the cumulative inventory difference test to detect a loss does not depend on the pattern of loss within the extended period but only upon the amount. The probability that at least one of the inventory differences included in a cumulative inventory difference will detect a loss does depend, among other factors, on how the loss is patterned over the periods $i = 1, 2, \dots, n$. If the amount M is to be diverted in n periods and the individual inventory differences have equal standard deviations, then the maximum probability of nondetection by at least one individual test occurs when a constant amount M/n is diverted in each of the n periods. If the amount M is diverted in one period, then the individual test for that period will have more loss detection power than a cumulative test that includes the period because an individual estimate will always have a smaller standard deviation than the cumulative estimate. Using the inventory difference and cumulative inventory difference test jointly guards against a multiplicity of alternative loss patterns. The individual tests guard against lumping all the loss into one period (abrupt diversion); cumulative tests guard against a gradual loss of the total amount over several periods, no matter what the specific loss pattern (protracted diversion).

Equations 16.9 and 16.11 describe the relationship between the four basic parameters of the testing procedure for either an inventory difference or a cumulative inventory difference. These are:

α : The probability of rejecting the null hypothesis when no loss or diversion occurs; the false alarm rate.

L_A, M_A : A loss or diversion which should be detected with reasonably high assurance; an amount of material of regulatory significance.

β : The probability of failing to reject the null hypothesis when the true loss is L_A or M_A .

$SE(\hat{ID}), SE(\hat{CID})$: The standard error of an estimate of the inventory difference or the cumulative inventory difference.

There are two basic ways in which this relationship is used. In the planning process, it is generally assumed that α , β , and L_A or M_A are determined by the objectives of the material accounting process. Given these three quantities, the $SE(\hat{ID})$ required to achieve these goals can be calculated. In essence, the quality of the measurement and accounting system necessary to achieve a given level of loss detection capability, as stated in terms of the parameters α , β , and L_A or M_A , has been established. Some of the ways in which these parameters can be related to the goals of the systems are considered in the next section.

When an evaluation is being performed after the fact, estimated values \hat{ID} and $\hat{SE}(\hat{ID})$ of the inventory difference and of its standard error are known from the calculation of the material balance and its variance from

available production and control measurements. From these observed values, assuming as before that the estimate is normally distributed, the probability α of observing an estimate as large as $\hat{I}D$ when $L = 0$, and the probability β of observing a value as small as $\hat{I}D$ when $L = L_A$, can be computed. Comparing these estimated values of α and β to the usual 0.05 or 0.01 significance levels is equivalent to testing whether L is significantly greater than zero or significantly less than L_A , respectively. Alternatively, using the estimated value of $\hat{I}D$ as the critical value for computing a value L_L of L associated with a given value of α and a value L_U of L associated with a given β is equivalent to determining a $100[1-(\alpha + \beta)]\%$ confidence interval for the true value of L . The determination of a confidence interval for L based on the estimate $\hat{I}D$ and its variance indicates the spread of possible true values of the inventory difference consistent with the observed estimates. The quantity $[SE(\hat{I}D)] [z_{1-\alpha} + z_{1-\beta}]$ measures the width of this confidence interval for a fixed value of $SE(\hat{I}D)$. Comparison of this width, which depends only on the standard error of the inventory difference estimate, with the loss L_A to be detected determines the ability of the system to detect a loss of this size. For the frequently used values $\alpha = \beta = 0.05$, $z_{1-\alpha} = z_{1-\beta} = 1.645$ and the estimated width of the confidence interval is 3.29 $[SE(\hat{I}D)]$. Equivalently, the largest value of $SE(\hat{I}D)$ for which a hypothesis test can be designed that is capable of detecting L_A with probability $1-\beta$ and with false alarm rate α is $SE(\hat{I}D) = L_A/3.29$.

16.2.2 Decision Theoretic Approaches

The interpretation of the observed inventory difference and its estimated standard error based on the simple quality control procedures of Section 4.7 and the hypothesis testing approach described in Section 4.1 were discussed in preceding section. The decision to reject the null hypothesis, or simply to investigate for the presence of assignable causes of lack of control, was based on the choice of a given multiple of the estimated standard error as a control limit. This multiple, in turn, was based on the choice of a significance level (or false alarm rate) and a distribution function for an inventory difference or cumulative inventory difference which enabled the translation of this selected probability into a critical value for determining the action to be taken. Because of the importance of decisions based on observed inventory differences and cumulative inventory differences, much work has been done to determine good methods for selecting a decision rule. This work will only be reviewed briefly in this volume. More complete expositions of such approaches to the determination of test procedures are available in the literature (Shipley, 1977, 1983; Avenhaus, 1977, 1986).

Statistical procedures for arriving at optimum decisions have been considered at three levels. The first level of approach, used in the previous

section, is based on the classical methods of hypothesis testing. For a given test statistic and assumed distribution, the power function describes the behavior of the test for all alternatives to the null hypothesis that are assumed possible. To derive tests which are in some sense optimum, the problem is formulated in terms of the possible decisions and states of nature. In Table 4.1, the possible decisions are "Reject H_0 " and "Do not reject H_0 "; the states of nature are " H_0 is true" and " H_0 is false." More generally, it is necessary to determine from a given set of observations and their assumed distribution the probability that any possible decision will be taken given a specific state of nature. For a fixed significance level, this enables a correct selection of test statistic and critical region by choosing the procedure that minimizes the number of incorrect decisions, but does not provide guidance as to the size of the critical region or, equivalently, the ratio of false alarm rate to the power of the test. When applied to the testing of a single inventory difference, procedures in this class involve only the selection of a one-sided or two-sided critical region based on the choice of admissible alternatives, but do not give guidance in the selection of appropriate significance levels.

The second and third levels of statistical decision procedures both involve the establishment of a loss table, or series of loss functions, describing in comparable units the consequences of each action given each state of nature. Again using Table 4.1 as an example, it is necessary to determine the costs associated with Type I and Type II errors, and either a zero cost or a gain (a negative cost) associated with each correct decision. Based on this concept of a loss function, a number of procedures have been developed which consider the decision problem as a game against nature with the loss table as a payoff matrix. In the simple case of Table 4.1, this means that α and β , the probabilities of Type I and Type II errors, respectively, are chosen so as to make the expected losses due to these two errors equal. In simple safeguards terms, the losses due to false alarms—i.e., unnecessary investigations, unwarranted plant shutdowns, or unjustified accusations of diversion or loss—should balance the losses due to undetected discards or diversions. Approaches of this type to the simple problem of interpreting a single observation of an inventory difference whose variance is known have been studied (Siri, Ruderman, and Dresher, 1978). Studies of this type can provide considerable insight into appropriate choices of false alarm rates but are highly dependent on the formulation of the payoff function. Similar approaches have been used to determine appropriate sampling procedures for the verification problems considered in Chapter 17.

At the third level of statistical decision procedures, prior distributions of the states of nature are also assumed, so that a single Bayesian risk can be calculated and the procedures designed to minimize this risk. In the simple case of one inventory difference, this is equivalent to assuming a

probability that a diversion will be attempted or a loss will occur (Bennett, 1970). Considerations of this type have been taken into account in the formulation of inspection approaches but not in the interpretation of observed material balances. The Neyman-Pearson hypothesis testing criteria and the game theoretic approach have been applied more extensively than the Bayesian approach (Avenhaus and Frick, 1974a,b; Shipley, 1983). An excellent summary of work on strategic analysis is given by Goldman (1984).

16.3 SPECIALIZED ESTIMATION AND TESTING PROCEDURES

16.3.1 Specialized Loss Estimators

The inventory difference and cumulative inventory difference estimates are based on straightforward mass balance considerations. Other specialized loss estimators have been developed because individual tests on each inventory difference do not use all the available mass balance accounting information in a sequence of consecutive inventory differences. For example, the estimate of the i^{th} inventory difference in a sequence of length n does not depend on or utilize any information from previous periods, and the cumulative inventory difference based on a sequence does not use intermediate inventory information. The following display indicates the data used and ignored in each instance.

| Loss estimate | Data used | Data ignored |
|-------------------|----------------------------------|--|
| \hat{ID}_i | I_{i-1}, T_i, I_i | $I_0; I_1, \dots, I_{i-2}; T_1, T_2, \dots, T_{i-1}$ |
| $C\hat{ID}_{(n)}$ | $I_0, T_1, T_2, \dots, T_n, I_n$ | I_1, I_2, \dots, I_{n-1} |

The estimate \hat{ID}_i also ignores the information $I_{i+1}, I_{i+2}, \dots, I_n, T_{i+1}, T_{i+2}, \dots, T_n$ generated subsequent to period i .

Loss estimators that can use all or part of the additional information require additional assumptions concerning the loss model; thus, the use of the term *specialized* to describe these estimates. They were developed to use the information from the material accounting process more completely and to create loss estimators that are more sensitive when the loss is assumed to be constant. The main technical criterion is that it be possible to represent the components of the material balance as linear combinations of loss and physical inventory parameters that reflect the hypothesized loss situation. When the more detailed loss models are valid, the sensitivity of the estimates increases relative to that of an inventory difference, or cumu-

lative inventory difference, with increasing n and with increasing λ , where λ is defined as the ratio of the physical inventory variance to the net transfer variance. Thus these specialized estimates provide useful complements to inventory differences and cumulative inventory differences when special loss situations are of concern and when λ and n are sufficiently large. The desirability of a large n suggests that sequences of inventory differences based on short and frequent accountability periods, thus approaching real-time accountability, are natural candidates for these procedures when the other conditions are satisfied.

The use or the contemplated use of a specialized loss estimator should involve the following considerations:

1. The detection of loss is an inherently important goal.
2. The loss models are not arbitrary but are based on strategic considerations in guarding against loss.
3. Relationships between the component variances such that a significant increase in loss detection power is possible are known to exist.
4. The overall Type I error rate can be controlled by simultaneous inference methods (Chapter 10).
5. An adequate computer program is available or feasible.
6. The loss estimators are obtained by regression techniques, so that the standard results associated with regression (Chapter 6) can be used to explore the adequacy of the model and/or the data.

Estimation procedures using generalized least squares are necessary when there are more equations than unknown parameters and the observations on the dependent variable are correlated and may have unequal variances. These conditions exist in specialized loss situations because in n consecutive periods there are $2n + 1$ observations ($n + 1$ inventory estimates and n net transfer estimates) and $n + 1 + p$ parameters (p loss parameters and $n + 1$ inventories) to be estimated (if $p < n$, then $n + 1 + p < 2n + 1$). The observed values of the inventories and net transfers are frequently correlated and usually have unequal variances. In the constant loss case the number of loss parameters is 1, the constant loss L per period.

In ordinary least squares, the dependent variables are uncorrelated with equal variances (Section 6.2.3). Under these conditions, the least squares estimates of the parameters are linear, unbiased, and have minimum variances. The theoretical basis for generalized least squares (Section 6.3.2) is that new variables that have equal variances and are uncorrelated are created by appropriate linear transformations, and the methods of ordinary least squares are then applied to these new variables. The variance-covariance matrix of the dependent variables is assumed to be known. In this application the estimated variances and covariances of the inventories and net transfers determined by the methods of Chapter 14

and 15 are used in the solution. There are special cases where other approaches can be used that are equivalent or approximately equivalent to generalized least squares estimates. These special cases occur when the loss model and/or the error model is simple.

In the remainder of this section a number of specific applications of specialized loss estimation are described briefly. For a more complete treatment the reader is referred to Bennett and Stewart (1983) and Wincek, Stewart, and Piepel (1979).

Case 1: The purpose here is to obtain and test the loss estimator for an abrupt, one-time loss in some one period by a process which proceeds sequentially. The hypothesis $H_0: L_1 = 0$ is tested first against the composite alternative hypothesis $H_1: L_1 > 0$ using the loss estimator $\hat{L}_1 = \hat{D}_1 = \hat{I}_0 + \hat{T}_1 - \hat{I}_1$.† If H_0 is not rejected, then the minimum variance linear unbiased estimator of μ_1 , the true beginning inventory of the second period, is $\hat{\mu}_1 = r_1 \hat{I}_1 + t_1 (\hat{I}_0 + \hat{T}_1)$, where r_1 and t_1 are obtained by minimizing the variance of $\hat{\mu}_1$ subject to the constraint $r_1 + t_1 = 1$. Then the loss estimator $\hat{L}_2 = \hat{\mu}_1 + \hat{T}_2 - \hat{I}_2$ is used to test $H_0: L_2 = 0$ against $H_1: L_2 > 0$. If H_0 is not rejected, $\hat{\mu}_2$ and \hat{L}_3 are obtained from $\hat{\mu}_2 = r_2 \hat{I}_2 + t_2 (\hat{\mu}_1 + \hat{T}_2)$ and $\hat{L}_3 = \hat{\mu}_2 + \hat{T}_3 - \hat{I}_3$, respectively, where r_2 and t_2 are obtained by minimizing the variance of $\hat{\mu}_2$ subject to the constraint $r_2 + t_2 = 1$. The process is continued until H_0 is rejected for some period.

This problem was first solved under the conditions $\text{Var}(\hat{I}_i) = \text{Var}(\hat{T}_i)$, $\text{Var}(\hat{T}_i) = \text{Var}(\hat{T}_j)$, and zero covariances (Stewart, 1958). The problem was generalized and solved for the case of zero covariances among the inventories and net transfers, but unrestricted variances (Stewart, 1972). This is an example of a relatively simple procedure that uses the information more completely and gives results that would be obtained under the generalized least squares approach. The problems of determining the operating characteristics of such sequential testing procedures are discussed by Wincek, Stewart, and Piepel (1979), Pike and Woods (1983), and Goldman (1985).

Case 2: The analyst wishes to use the data more completely to test for an abrupt, one-time loss in a selected intermediate period i out of a sequence of n periods. The estimate of the beginning inventory μ_{i-1} for period i is obtained by the procedure of Case 1. The ending inventory value μ_i for period i is estimated as follows. Estimate μ_{n-1} by $\hat{\nu}_{n-1} = r_{n-1} \hat{I}_{n-1} + t_{n-1} (\hat{I}_n - \hat{T}_n)$, where the values r_{n-1} and t_{n-1} are obtained by minimizing the variance of $\hat{\nu}_{n-1}$ subject to the constraint $r_{n-1} + t_{n-1} = 1$. Then estimate μ_{n-2} by $\hat{\nu}_{n-2} = r_{n-2} \hat{I}_{n-2} + t_{n-2} (\hat{\nu}_{n-1} - \hat{T}_{n-1})$. This process is continued until $\hat{\nu}_i$ is obtained. Then the loss estimator $\hat{L}_i = \hat{\mu}_{i-1} + T_i - \hat{\nu}_i$ is used to test $H_0: L_i = 0$ against $H_1: L_i > 0$. This problem is discussed by Wincek, Stewart, and Piepel (1979).

† \hat{D}_i is used to represent the observed inventory difference for the i^{th} period in the remainder of this section.

Case 3: In this case, the losses are assumed to be constant from period to period and a constant loss estimator is used. The constant loss estimator uses a more specific loss model than the cumulative inventory difference and hence, uses the information available for the sequence of periods more completely. The constant loss estimator was first developed for the same variance conditions as in Case 1. The methods that have been derived to obtain $\hat{L}_{(n)}$, the estimate of L based on information from n successive periods, include: (1) Kalman Filter techniques (Pike and Morrison, 1977); (2) minimizing the variance of linear combinations of D_i values (Jaech, 1978); (3) the use of orthogonal linear forms (Wincek, Stewart, and Piepel, 1979); and (4) generalized least squares (Wincek, Stewart, and Piepel, 1979). These approaches, which seem to be so different, all provide identical minimum variance linear unbiased estimators of L .

Case 4: In this case, the constant loss problem is generalized to the situation where nonzero covariances exist among the material balance components and the variances are unrestricted. For the constant loss model, only the inventory differences and their variance-covariance matrix are needed to obtain the least squares estimate of L based on n periods. Because $E(\hat{D}_i) = L$, the matrix representation of the model is

$$\underline{D} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} (L) + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \underline{X}L + \underline{\epsilon} \quad (16.12)$$

so that by the generalized least squares procedures of Section 6.3.2

$$\hat{L}_{(n)} = (\underline{X}' \underline{\Sigma}_{\epsilon}^{-1} \underline{X})^{-1} \underline{X}' \underline{\Sigma}_{\epsilon}^{-1} \underline{D} \quad (16.13)$$

and

$$\text{Var}(\hat{L}_{(n)}) = (\underline{X}' \underline{\Sigma}_{\epsilon}^{-1} \underline{X})^{-1} \quad (16.14)$$

Because $\underline{X}' = (1, 1, \dots, 1)$ is a $1 \times n$ unit vector, the variance of $\hat{L}_{(n)}$ can be written

$$\text{Var}(\hat{L}_{(n)}) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n \sigma_{(n)}^{ij}} \quad (16.15)$$

where $\sigma_{(n)}^{ij}$ is the i^{th} row, j^{th} column element of the $(n \times n)$ matrix $\underline{\Sigma}_{\epsilon}^{-1}$.

Case 5: The analyst wishes to estimate the total loss for n periods under the hypothesis $H_0: L_i = \gamma_i M$, where $\sum_{i=1}^n \gamma_i = 1$ and γ_i is the proportion of material that is lost in period i , and M is the total loss for the n periods. Here γ_i is a design variable and M is a parameter. Then the inventory differences are modeled as

$$\underline{D} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} (M) + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \underline{\Gamma} M + \underline{\epsilon} \quad (16.16)$$

and it can be shown that \hat{M} , the minimum variance linear unbiased estimate of M , is

$$\hat{M} = \frac{\sum_{i=1}^n \gamma_i \hat{D}_i \sum_{j=1}^n \gamma_j \sigma_{(n)}^{jj}}{\sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j \sigma_{(n)}^{jj}} \quad (16.17)$$

with

$$\text{Var}(\hat{M}) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j \sigma_{(n)}^{jj}} \quad (16.18)$$

This formulation of the loss problem is useful in guarding against a particular pattern of loss. For example, the γ values may be made proportional to period throughputs.

Case 6: More than one loss parameter can be used. Suppose the analyst wants to estimate and to test for a constant loss in Periods 2, 3, and 4 and an abrupt loss in Period 6 in a sequence of seven material balances. The inventory differences could be modeled as

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \\ D_6 \\ D_7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \end{bmatrix}$$

The value \hat{L}_1 is an estimate of the constant loss value for Periods 2, 3, and 4; and \hat{L}_2 is an estimate of the one-time loss in Period 6.

These different possibilities illustrate: (1) some of the many loss models that are possible; (2) the generality of the approach to these problems and modifications of the general approach that are helpful in simplifying the computations; and (3) some special properties of constant loss estimators.

16.3.2 Sequential Tests

The examples of the preceding section emphasize the estimation of loss based on the data for a sequence of n material balance periods. These estimation procedures are characteristic of procedures for nuclear materials accounting and management, where the primary requirement is the assessment of the status of the nuclear material. However, the shift to detection of loss or deliberate diversion as the primary function of the material accounting process, and the consequent need for frequent material balances as a means of obtaining timely decisions that diversion has occurred, changed the emphasis to decisions as opposed to assessment (Pike, Woods, and Rose, 1980; Shipley, 1981, 1983). Numerous procedures have been proposed for testing material balance data under these new conditions (Stewart, 1978; Jaech, 1978; Avenhaus and Frick, 1979; Avenhaus and Beedgen, 1980; Cobb et al., 1981), particularly in connection with extensive studies of real-time accounting. Reviews of these developments (Shipley, 1981; Goldman, Picard, and Shipley, 1982; Gupta, 1983; Pike and Woods, 1983) emphasize three aspects of the problem:

1. The need to develop tests, or groups of tests, which are robust against all possible diversion tactics.
2. The need to develop tests that are truly sequential, rather than simply based on more than one material balance, so as to emphasize timely detection.
3. The need to characterize test performance in at least the minimum decision-theoretic sense of being able to place upper bounds on detection probability and lower bounds on false alarm rates.

The multiplicity of tests involved in both sequential test procedures and multiple tests makes it difficult to satisfy these three requirements simultaneously. Shipley (1981, 1983) has stated the criteria for a satisfactory statistical testing procedure as follows:

1. To preserve timeliness, the opportunity to decide that a loss has occurred must be available soon after each balance.
2. Losses of sufficient size occurring over *any* time interval, up to one-year long, for example, must be detectable. (One year is chosen because that seems to be the currently accepted standard for the maximum interval.)

3. A decision that diversion has or has not occurred during a time interval, and backed up by appropriate assessment procedures, may not be changed at a later time in the absence of substantial new information.

4. A definitive estimate of the inventory difference and its variance must be available periodically, say, once each calendar year at physical inventory time.

5. The false alarm and detection probabilities and the detection sensitivity (goal quantity) must be known (or at least bounded) beforehand.

If one adds to these the requirement that the procedures can be applied and understood in connection with routine evaluation in the field, it is easy to see why development and acceptance of appropriate tests has been difficult.

With respect to the requirement to detect all types of loss patterns, four procedures involving sequences of material balances seem to cover most cases. The first is the cumulative inventory difference, which is a specific case of the cumulative sum, or CUSUM, frequently used as a statistic in sequential testing (Section 4.6). The cumulative inventory difference is attractive as a test statistic because its power, as noted earlier, is independent of the loss pattern, depending only on the total loss $M = \sum L_i$. The second is the uniform loss estimator of Cases 3 and 4 of the previous section. As noted there, this case has received much attention and is extremely sensitive to the early detection of protracted diversion which is relatively uniform (e.g., Case 5 of the previous section, where the loss is proportional to the throughput). The third is the procedure of Case 1 of the preceding section, which is sensitive to change in the loss pattern (Wincek, Stewart, and Piepel, 1979; Sellinschegg et al., 1981). This procedure detects the initiation of diversion or loss, as opposed to the presence of a protracted loss. The fourth procedure is a test sensitive to the variance in the losses which will protect against deliberate randomization of individual losses to counter the sensitivity of the estimators based on constant loss.

Each of the above estimates can be computed at the time of the n^{th} material balance using any number 1 to n of consecutive prior balances, and its significance tested against its estimated standard error. Interpreting the very large number of test statistics so obtained is difficult. Computerized display procedures which aid in this process have been developed (Shipley, 1978). However, they do not quantify the performance of these highly dependent multiple tests with respect to their detection capability. For independent sequential statistics derived from a sequence of inventory differences, sequential probability ratio tests (Section 4.6), Page's test, and tests of power one have been studied and their properties determined (Goldman, 1985).

Relatively little plant experience with sequential procedures is available. Demonstrations have been carried out in several large reprocessing plants (Cobb et al., 1981; Ikawa, et al., 1983).

REFERENCES

- Avenhaus, R., 1977, *Material Accountability—Theory Verification, Applications*, Monograph of the Wiley IIASA International Series on Applied Systems Analysis, John Wiley & Sons, Inc., Chichester, England.
- , 1986, *Safeguards Systems Analysis*, Plenum Press, New York.
- , and R. Beedgen, 1980, Statistical Analysis of Alternative Data Evaluation Schemes, Part II, *Proc. 2nd ESARDA Annu. Symp. on Safeguards and Nucl. Mater. Manage.*, Edinburgh, 167-172.
- , and H. Frick, 1974a, *Game Theoretical Treatment of Material Accountability Problems*, International Institute for Applied Systems Analysis, RR-74-2. A slightly modified version has been published under the same title in *Int. J. Game Theory*, 1977, 5: 117-135.
- , and H. Frick, 1974b, *Game Theoretical Treatment of Material Accountability Problems, Part II*, International Institute for Applied Systems Analysis, RR-74-21.
- , and H. Frick, 1979, Statistical Analysis of Alternative Data Evaluation Schemes (Part I), *Proc. 1st ESARDA Symp.*, Brussels, 442-446.
- Bennett, C. A., 1970, Progress in Systems Analysis, *Safeguards Techniques*, II: 247-254, International Atomic Energy Agency, Vienna.
- , and K. B. Stewart, 1983, Efficient Estimation of Material Loss, *Mathematical and Statistical Methods in Nuclear Safeguards*, Harvard Academic Publishers, Boston.
- Cobb, D. D., H. A. Dayem, A. L. Baker, J. H. Ellis, M. H. Ehinger, and J. M. Crawford, 1981, Demonstration of Near-Real-Time Accounting: The AGNS 1980 Miniruns, *Nucl. Mater. Manage.*, X(1): 34-43.
- Fienburg, S. E., J. G. Neter, and K. A. Leitch, 1977, Estimating the Total Overstatement Error in Accounting Populations, *J. Amer. Stat. Assoc.*, 72: 295-302.
- Goldman, A. J., 1984, *Strategic Analysis for Safeguards Systems: A Feasibility Study*, NUREG/CR-3926, Vol. 2. U. S. Nuclear Regulatory Commission, Washington, D. C. Available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- Goldman, A. S., 1985, *Sequential Procedures for Detecting Material Losses*, NUREG/CR-4107, U. S. Nuclear Regulatory Commission, Washington, D. C. Available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- , R. R. Picard, and J. P. Shipley, 1982, Statistical Methods for Nuclear Materials Safeguards: An Overview, *Technometrics*, 24: 267-274.
- Gupta, D. (Ed.), 1983, *Overview Report of the International Workshop on the Near-Real-Time Accountancy Measure*, Report No. KFK-3515, Karlsruhe Research Center, FRG.
- Ikawa, K., H. Ihara, H. Nishimura, M. Tsutsumi, and T. Sawahata, 1983, A Near-Real-Time Material Accountancy Model and Its Preliminary Demonstration in the Tokai Reprocessing Plant, *Proc. International Symposium, Vienna, 1982*, International Atomic Energy Agency, Vienna, Vol. I, pp. 499-512.
- Jaech, J. L., 1978, On Forming Linear Combinations of Accounting Data To Detect Constant Small Losses, *Nucl. Mater. Manage.*, VI(4): 37-41.
- , and C. G. Hough, 1983, Consequences of Discrepancies on Verified Material Balances, *Proceedings of a Symposium, Vienna, November, 1983*, International Atomic Energy Agency, Vienna, Vol. II, pp. 287-304.
- Neter, J., and J. R. Loebhecke, 1977, On the Behavior of Statistical Estimation When Sampling Accounting Populations, *J. Am. Stat. Assoc.*, 72: 501-507.
- Pike, D. J., and G. W. Morrison, 1977, A New Approach to Safeguards Accounting, *Nucl. Mater. Manage.*, VI(3): 641-658.
- , and A. J. Woods, 1983, Statistical Methods in Nuclear Material Accountancy, *Proc. Symposium, Vienna, November, 1982*, International Atomic Energy Agency, Vienna, II: 359-372.
- , A. J. Woods, and D. M. Rose, 1980, *A Critical Appraisal of the Use of Estimators for the Detection of Loss in Nuclear Material Accountancy*, University of Reading Technical Report 1/80/07, Reading, England.

- Sellinschegg, D., U. Bicking, K. H. Cerff, and G. Naegele, 1981, Dynamic Material Accountancy in a Large Reprocessing Facility, *Proc. 3rd Annu. Symp. on Safeguards and Nucl. Mater. Manage.*, ESARDA, 13: 323-328.
- Shipley, J. P., 1977, Decision Analysis in Safeguarding Special Nuclear Material, *Trans. Am. Nucl. Soc.*, 27: 178.
- , 1978, Efficient Analysis of Dynamic Materials Accounting Data, Proc. 19th Annu. Meeting Inst. Nuc. Mater. Manage., Cincinnati, Ohio, 1978, *Nucl. Mater. Manage.*, VII: 355-366.
- , 1981, Decision-Directed Materials Accounting Procedures: An Overview, Proc. 22nd Annu. Meeting, Inst. Nucl. Mater. Manage., San Francisco, California, 1981, *Nucl. Mater. Manage.*, X: 281-287.
- , 1983, Analysis of Materials Accounting Data, *Mathematical and Statistical Methods in Nuclear Safeguards*, Harvard Academic Publishers, Boston.
- Siri, W., H. Ruderman, and M. Dresner, 1978, *The Application of Game Theory to Nuclear Material Accounting*, NUREG/CR-0490, U. S. Nuclear Regulatory Commission, Washington, D. C.
- Stewart, K. B., 1958, *B-PID and Inventory Estimates with Minimum Variance*, HW-56536, Hanford Laboratories Operation, U. S. Atomic Energy Commission, Washington, D. C., Available from NTIS, U. S. Department of Commerce, Springfield, Virginia.
- , 1972, Situations in Which the Arithmetic and Geometric Moving Averages Are Inefficient Estimators, *Technometrics*, 14: 619-633.
- , 1978, The Loss Detection Powers of Four Loss Estimators, *Nucl. Mater. Manage.*, VII(3): 74-80.
- Wincek, M. A., K. B. Stewart, and G. F. Piepel, 1979, *Statistical Methods for Evaluating Sequential Material Balance Data*, NUREG/CR-0683 (PNL-2920), Pacific Northwest Laboratory, Richland, Washington.

CHAPTER 17

Verification of Accounting Data

17.0 INTRODUCTION

In Chapter 13, the basic concepts of nuclear material accounting were introduced. The many close parallels between the concepts of material accounting and those of standard financial accounting were highlighted. In this chapter, the verification of material accounting data is treated. These verification activities are parallel to the audit procedures of financial accounting.

This chapter consists of three main sections. Section 17.1 presents an overview of the verification process in which the statistical aspects of the problem are outlined. Section 17.2 treats statistical methods for making inferences about the "errors" or "discrepancies" (these terms will be defined later) that may be uncovered in the course of verifying accounting data. Section 17.3 presents a statistical method for verifying the accounting results. Specifically, it provides a procedure for making inferences concerning the correctness of reports of amounts of material shipped, received, or on hand, as well as the material balance achieved.

17.1 THE VERIFICATION PROCESS

To provide a background for discussion of specific verification techniques useful in nuclear material accounting applications, a brief overview of the verification process is presented and the central concepts are introduced. As used in this chapter, **verification** refers to that part of the audit process that concentrates primarily on the *independent collection, examination, and evaluation of data* in order to verify the correctness of the reported accounting results. It is the aspect of auditing in which statistical methods are most commonly used.

Auditing has been defined as "a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users" (Committee on Basic Auditing Concepts, 1972). In the specific context of nuclear material safeguards, ANSI Standard N15.38-1982 (ANSI, 1982) defines an audit as "a documented examination, evaluation, and correction activity, performed in accordance with this standard, to verify that all or a portion of a safeguards system or subsystem has been ade-

quately developed, documented, and implemented as necessary to satisfy applicable requirements." In the United States, the applicable requirements are given, for example, as:

1. Atomic Energy Act of 1954, as amended.
2. Energy Reorganization Act of 1974, as amended.
3. Code of Federal Regulations.
4. Department of Energy orders.
5. Operating licenses, contracts, and conditions.
6. Regulatory guides.
7. Operating procedures and practices.
8. American National Standards.

Different requirements may apply in other countries. It should be noted that generally the tasks of auditing, and verification in particular, are strongly influenced by the nature of the criteria established to govern the nuclear material accounting system. For example, the ground rules and constraints for verification activities are very different in international safeguards and domestic safeguards, although many of the same statistical techniques can be used.

The first definition of auditing given above emphasizes the concept of evidence as the basis for audit conclusions. In ANSI Standard N15.38-1982, the role of objective evidence is also stressed (ANSI, 1982):

Sufficient valid and relevant evidential matter shall be obtained and made part of the working papers through inspection, observation, inquiries, and confirmations to afford a reasonable basis for the report regarding the compliance of the nuclear materials safeguards system under investigation.

The verification process concentrates on a particular kind of evidence—namely, quantitative data—and this is where statistics enters the picture. The role of statistics is largely directed toward objectively obtaining and evaluating sufficient quantitative evidence. Statistical sampling plans are used to obtain such evidence objectively. Statistical estimation procedures and hypothesis tests are used to evaluate the evidence objectively. Statistical design criteria for sampling and evaluation are used to ensure that the evidence will be sufficient.

17.1.1 The Internal Control System

In the financial context, all but the very simplest accounting systems make provisions for an internal control system that establishes and enforces uniform accounting procedures and monitors the quality of the information generated by the accounting system. In U.S. domestic safeguards, the counterpart of the internal control system is the material control and accounting (MC&A) system. For international (IAEA) safeguards, it is the State System of Accounting and Control (SSAC). In each case, this internal control system is of fundamental importance in the

auditing and verification process. The nature and quality of the internal control system largely determine the required scope and intensity of the audit and verification activities, and furthermore have an important impact on the level of assurance that can be derived from auditing and verification. An excellent internal control system not only facilitates auditing and verification, but also strengthens the conclusions that can be drawn. A weak or ineffective system, on the other hand, may require an inordinate amount of effort to audit and verify, and, even in the most favorable circumstances, may permit only a qualified statement or conclusion about the accounts being examined.

Auditing is a multistage process in which the internal control system plays a central role. Among the determinations that must be made are the following:

1. Is the **design** of the internal control system adequate to prevent and/or control the occurrence of errors, intentional or unintentional?
2. Has the **implementation** of the internal control system been adequate to ensure that the system functions as it was designed to function, thereby providing some assurance of control over possible errors?
3. Do **direct tests** of the accounts, performed independently by the auditor, substantiate the effectiveness of the internal control system as implemented?

Quantitative data and statistical methods do not play a major role in helping the auditor address the first question. The adequacy of an internal control system design depends on such matters as the establishment of clear lines of authority and responsibility, the existence of documented standard procedures (some of which may involve statistical considerations), the allocation of sufficient resources to the accounting and control functions, segregation of functional responsibilities, and staffing and training policies. In addressing the second and third areas, however, quantitative data and statistical methods have a more significant role. The verification procedures described in this chapter are aimed at these latter areas. The quality of implementation of the internal control system can be evaluated through **compliance tests**† performed on a statistical sample of transactions and procedures. The overall effectiveness of the internal control system can be substantiated or confirmed through **substantive tests**† of the accounts, independently carried out by the auditor, again on a statistical basis.

In the safeguards literature, using a somewhat different nomenclature, the three levels or stages of auditing/verification outlined above are sometimes referred to as: (1) capability, (2) performance, and (3) independent assurance. This usage is illustrated in the following extract from the

† These concepts will be explained and illustrated more fully in the rest of the chapter.

IAEA's Guidelines for States' Systems of Accounting for and Control of Nuclear Materials (IAEA, 1980):

Assurance of operator compliance with the requirements of the system of accounting and control established by the State and the assessment of its effectiveness can be achieved only by means of a comprehensive audit and inspection programme. Such a programme should have the first of the following objectives, and may have others, including the second:

(a) To ensure that the capability of, and performance by, each facility operator for the discharge of his responsibility for accounting for and control of nuclear material satisfy the requirements of the Agreement with the IAEA.†

(b) In addition, a State may wish to derive assurance, through independent verification at facilities by the Authority, that the accounting and control measures implemented by the facility operator are effective and, in conjunction with other measures, to conclude that there has been no unauthorized removal or use of nuclear material.

These assurances contribute to the establishment by the IAEA, through its independent verification activities, whether there has been any diversion of significant quantities of nuclear material. The Authority should establish criteria against which the operator's capability and performance and the results of SSAC inspections and evaluations can be assessed.

Although this terminology is somewhat different from that introduced earlier, the underlying concepts are identical.

17.1.2 Fundamental Concepts

The essence of verification is the independent collection, examination, and evaluation of data to (1) ensure that the internal control system is functioning as designed, thereby obtaining some assurance of control over possible errors, and (2) substantiate the quantities of material recorded by the accounting system. In the first instance, the emphasis is on characterizing and evaluating the errors in the accounts. In the second, the emphasis is on the material itself and, ultimately, on whether some material might have been lost or diverted or might otherwise be unaccounted for.

In most cases, verification involves the comparison of two independently generated sets of data. The principal examples are inspection data and shipper-receiver data. During an inspection, the inspection team obtains independent measurements of a sample of items. (Recall from Chapter 13 the definition of an item, the primary accounting unit in a material accounting system, as an object or container that has a unique material quantity assigned to it and which can be identified and accounted for as a separate entity.) These inspection measurements can be compared with the data recorded by the plant operator for the same items. Ship-

† In this statement, "capability" refers to the actual presence at the facility of the records, equipment, trained personnel, documented procedures, and other resources needed for adequate accounting and control of nuclear material at the facility, and "performance" refers to the actual use of the capability for accounting and control of nuclear material at the facility. Objective (a) may include some independent verification activities by the Authority (e.g., through measurements and observations).

per-receiver data consist of independent measurements made by both the shipper and receiver on each item in one or more shipments of nuclear material. From a statistical viewpoint these two situations are similar. Both involve paired data, and the same statistical techniques are generally applicable. The discussion in this chapter will focus primarily on inspection data, i.e., operator-inspector comparisons. However, on occasion, reference will also be made to the shipper-receiver context.

Generally, the inspector will make two different kinds of measurements which require different kinds of statistical procedures for comparing operator and inspector data. The two types are referred to in safeguards parlance as attributes and variables measurements. These terms are derived from quality control conventions. An **attributes measurement** or test simply classifies an item as either "defective" or "not defective," depending on whether or not the item possesses some specified attribute or characteristic. In the language of Chapter 1, it is a measurement on a (dichotomous) nominal scale. In verification applications, the attribute of interest is usually the correctness of the item amount stated by the plant operator. More generally, the attribute may be the correctness of any assertion or implication concerning the accounting system. A **variables measurement** is a measurement, on a continuous scale, of some characteristic or quality an item possesses. In nuclear material accounting, examples of variables measurements include masses and weights, volumes, concentrations and enrichments. In the language of Chapter 1, a variables measurement is a measurement on a ratio scale.

Errors in nuclear material accounting data can be classified in several ways. One important classification is in terms of the source or cause of the error. One source that is always present is measurement variability. This source of error can usually be characterized in terms of one or more random variables. Ways to estimate and control measurement error were discussed in Chapters 14 and 15. A second broad category of errors involves the bookkeeping or clerical mistakes which occur in any accounting system. Examples include mistakes in transcribing data from one record to another, mistakes in calculations, and mistakes in applying prescribed procedures. These mistakes vary widely both in magnitude and in frequency of occurrence and are difficult to characterize statistically, but they are unavoidable in practice. An effective internal control system can minimize but never completely eliminate them. A third distinct possible cause of error is the willful manipulation of the accounting system to conceal a diversion or loss of material. Again, an effective internal control system can minimize the possibility that such discrepancies will go undetected but cannot completely rule them out.

A second important classification of errors is based on the size of the effect or discrepancy. The definition of size depends on the situation and well-defined rules are not possible. However, in safeguards applications it is common to base the definition on the detection sensitivities of the verifi-

cation techniques used in inspection. Generally, these methods are of two kinds:

1. **Nondestructive assays (NDA).** These are usually relatively rapid instrumental techniques.
2. **Destructive analyses.** These are relatively more time-consuming, with higher precision than the NDA techniques. They normally involve separate bulk measurements (weight or volume), a bulk sampling operation, and elemental and/or isotopic analysis.

These two methods of verification furnish the basis for defining three error size classifications:

1. **Gross defects.** A gross defect is an error that is large enough to be detected with essentially 100% certainty by a single nondestructive measurement. Thus, the size of the defect must be large compared with the variability in the nondestructive measurement.
2. **Medium defects** (sometimes called partial defects). A medium defect can be detected with essentially 100% certainty by a single measurement by a destructive analysis technique. Such a defect must be large compared with the variability in the destructive analysis but not large enough to be detected with 100% certainty by NDA.
3. **Small defects** (sometimes called biases). A small defect is an error that is too small to be detected by a single measurement of either type. However, the cumulative effect of a number of small defects may be detectable by statistical tests performed on a set of data generated by verifying a collection of items.

A third important concept in characterizing the size of an error is the concept of a **significant quantity**, variously referred to as target quantity, strategic goal quantity, formula quantity, or threshold amount. The analogous concept in financial auditing is materiality. Whatever the terminology, the essential point is that there is a certain size or amount of total discrepancy or loss that is regarded as practically significant in some well-defined sense; in this context "significant" does not mean the same thing as "statistically significant." In safeguards applications, the significant quantity is usually defined as a matter of regulatory or administrative policy.

The relevance of this concept to verification is that the intensity of verification is usually governed by the desire to detect, if missing, the loss or diversion of a significant quantity or formula quantity. Thus, this quantity enters into the determination of sample sizes for inspection. The size of an observed discrepancy relative to the significant or formula quantity also determines to some extent the nature and intensity of the follow-up activities prescribed when errors are uncovered. For obvious reasons, discrepancies of little or no practical significance will not generally warrant the same degree of attention and concern as those that are large relative to the significant quantity.

17.1.3 Selecting a Random Sample

The process of collecting verification data has several statistical aspects. First, items to be verified must be selected on a statistical (random sampling) basis. Second, measurements of the selected items must be performed by measurement systems that are in a state of statistical control. Statistical sampling techniques are presented in Chapter 8. Problems of measurement control have been covered in Chapter 15 and the techniques presented there are equally applicable to inspection measurements and facility measurements. Specific procedures for selecting a random sample in this context have not been discussed in other chapters, so an account of the process is given here.

There are three requirements for selecting a random sample: the **sampling frame**, the **sample size**, and a **source of random numbers**. The frame is usually thought of as a list of all items in the population (universe) to be verified. A more general definition is given by Deming (1960): "The frame is a set of physical materials (census statistics, maps, lists, directories, records) that enables us to take hold of the universe piece by piece." The more general concept includes physical lists and also procedures that can account for all the items in the population without actually going to the trouble of listing them. In nuclear material accounting, the frame is commonly an inventory listing. It may also be a list of items on a shipping document, a list of storage locations in a critical assembly, a list of transactions in a journal, and so forth. It is worth noting that the frame may be, and often is, in error. The list may not be a completely accurate representation of the population to be verified. Some items in the population may fail to appear in the list. Other items may inadvertently appear twice. Items on the list may be mislabeled or misidentified in any number of ways. One of the purposes of verification is to check for deficiencies in the frame.

The frame is used as the basis for sample selection. Given a sample size determined in accordance with a statistical plan, random numbers are used to select items in the frame for verification. The approaches that can be used are described in Section 11.3.1. The simplest and most widely used in verification of nuclear material accounting results are based on tables of random numbers or simple random number generating routines contained in hand-held calculators. Table 17.1, taken from Snedecor and Cochran (1980), gives an example of 1000 random digits grouped in blocks of five. Many statistical texts and reference volumes of statistical tables include a collection of random numbers which are adequate for the selection practices described here. The routines used to generate random numbers using small computers are also generally adequate for sample selection, although they may not be satisfactory in situations where large numbers are required, such as the simulation techniques described in Chapter 11.

There are many ways to choose random numbers for sample selection that work equally well. For example, any method of stepping through the

TABLE 17.1
One Thousand Random Digits†

| | 00-04 | 05-09 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 00 | 54463 | 22662 | 65905 | 70639 | 79365 | 67382 | 29085 | 69831 | 47058 | 08186 |
| 01 | 15389 | 85205 | 18850 | 39226 | 42249 | 90669 | 96325 | 23248 | 60933 | 26927 |
| 02 | 85941 | 40756 | 82414 | 02015 | 13858 | 78030 | 16269 | 65978 | 01385 | 15345 |
| 03 | 61149 | 69440 | 11286 | 88218 | 58925 | 03638 | 52862 | 62733 | 33451 | 77455 |
| 04 | 05219 | 81619 | 10651 | 67079 | 92511 | 59888 | 84502 | 72095 | 83463 | 75577 |
| 05 | 41417 | 98326 | 87719 | 92294 | 46614 | 50948 | 64886 | 20002 | 97365 | 30976 |
| 06 | 28357 | 94070 | 20652 | 35774 | 16249 | 75019 | 21145 | 05217 | 47286 | 76305 |
| 07 | 17783 | 00015 | 10806 | 83091 | 91530 | 36466 | 39981 | 62481 | 49177 | 75779 |
| 08 | 40950 | 84820 | 29881 | 85966 | 62800 | 70326 | 84740 | 62660 | 77379 | 90279 |
| 09 | 82995 | 64157 | 66164 | 41180 | 10089 | 41757 | 78258 | 96488 | 88629 | 37231 |
| 10 | 96754 | 17676 | 55659 | 44105 | 47361 | 34833 | 86679 | 23930 | 53249 | 27083 |
| 11 | 34357 | 88040 | 53364 | 71726 | 45690 | 66334 | 60332 | 22554 | 90600 | 71113 |
| 12 | 06318 | 37403 | 49927 | 57715 | 50423 | 67372 | 63116 | 48888 | 21505 | 80182 |
| 13 | 62111 | 52820 | 07243 | 79931 | 89292 | 84767 | 85693 | 73947 | 22278 | 11551 |
| 14 | 47534 | 09243 | 67879 | 00544 | 23410 | 12740 | 02540 | 54440 | 32949 | 13491 |
| 15 | 98614 | 75993 | 84460 | 62846 | 59844 | 14922 | 48730 | 73443 | 48167 | 34770 |
| 16 | 24856 | 03648 | 44898 | 09351 | 98795 | 18644 | 39765 | 71058 | 90368 | 44104 |
| 17 | 96887 | 12479 | 80621 | 66223 | 86085 | 78285 | 02432 | 53342 | 42846 | 94771 |
| 18 | 90801 | 21472 | 42815 | 77408 | 37390 | 76766 | 52615 | 32141 | 30268 | 18106 |
| 19 | 55165 | 77312 | 83666 | 36028 | 28420 | 70219 | 81369 | 41943 | 47366 | 41067 |

†Reprinted with permission from Snedecor and Cochran (1980).

numbers in the table is equally valid—vertically down the columns, horizontally across the rows, diagonally—because the numbers in the table are completely random. For the sake of convenience, it is usually desirable to pick a simple systematic method, such as searching down columns. The choice of entry point or starting point is completely arbitrary, again because the numbers are random, but a different starting point should be selected each time the table is used. A simple way to accomplish this is to make a pencil mark after the last random number used in a particular sampling plan. The next time a sample is drawn, simply start at that point.

The use of the tables can be inefficient if the size of the population is substantially smaller than a power of 10. For example, suppose the population to be verified contains 15 items. Only about 15% of the two digit numbers in the table will be between 1 and 15, with the result that most of the numbers in the table will simply be rejected and it may take a little longer to find the sample. There are simple modifications whereby this inefficiency can be avoided. For example, subtract 20 from all numbers between 21 and 40, subtract 40 from numbers between 41 and 60, etc. In

this way, every two digit random number will give rise to a random number between 1 and 20, and about 75% of these will be between 1 and 15. Thus, the rejection rate will be much lower than before. Note that the number subtracted must be an exact divisor of the appropriate power of 10 used to enter the table.

Example 17.1 The inspector wishes to verify a stratum consisting of 48 cylinders of uranium hexafluoride (UF_6). Calculations of the required sample size indicate that 7 of the 48 cylinders should be selected for inspection. First, the UF_6 cylinders are numbered sequentially from 1 to 48. This is easily done using the inventory list. Then an arbitrary starting point or entry point for entering the random number table is selected, and the following sequence of two-digit numbers is systematically searched for numbers between 1 and 48. For example, suppose he starts with row 5, columns 18 and 19, and moves down the columns. The first number is 94, the second 74, the third 91, the fourth 66, the fifth 80, the sixth (row 10, columns 18 and 19) is 05. This is between 1 and 48, so the inspector puts cylinder number 5 in the sample. The next 9 two-digit numbers in columns 18 and 19 are 26, 15, 31, 44, 46, 51, 23, 08, and 28. All except 51 are between 1 and 48, so the inspector includes in the sample cylinders 26, 15, 31, 44, 46, and 23, at which point the desired sample size of 7 has been reached.

Unless the sample size is large, using a computer program to generate random numbers is not likely to be a great deal faster than using a random number table. However, computerized methods have the important advantage of being self-documenting. The sampling table printed out by the computer can be included with the inspection working papers or reports as a record of what was done.

17.2 INFERENCES FROM OBSERVATION OF ATTRIBUTES

One of the purposes of verifying accounting data is to obtain information about the existence of mistakes or discrepancies in the reported data. Such information can be used to indicate:

1. Whether the implementation of the internal control system has been adequate. Ideally, an effective internal control system will permit only a small number of errors in the data, none of which will seriously distort the accounting results.

2. Whether a loss or diversion of material might have occurred. Many possible scenarios by which loss or diversion could be concealed will manifest themselves in terms of discrepancies in the accounting data.

Most of the statistical techniques to be presented in this section are dual-purpose in the sense that they provide evidence concerning both of the possibilities mentioned above. It is useful to keep these two possibilities firmly in mind even though the same statistical techniques are applied in assessing them. The distinction is important in interpreting the results of the statistical procedures and in defining the necessary follow-up actions.

The basic features of verification procedures based on the observation of attributes are described in the next section. Section 17.2.2 explains the use of the hypergeometric distribution in attribute sampling. Section 17.2.3 presents the use of the binomial distribution.

17.2.1 Attribute Sampling

A population of items is given and the inspector wishes to determine the proportion of items in the population that possess some specific attribute. To do this, the inspector selects a random sample of items from the population and subjects each selected item to a test (measurement, examination) to determine whether the item possesses the specified attribute or characteristic. Items that do not are usually called **defective**, e.g., containers with a broken seal. After examining each sampled item, the inspector can calculate the proportion of defectives in the sample and on the basis of this information can reach a conclusion about the proportion of defectives in the population as a whole. In verification applications, the attribute of interest is usually the correctness of the item amount as stated by the plant operator. However, correctness may also refer in some cases to compliance with prescribed procedures (i.e., compliance with the internal control system) regardless of whether the item amount is correctly stated. For example, if source documents are not properly signed and dated or if forms are not correctly filled out, the inspector could classify the inspected item as defective even if the item amount is correct.

In practice, the definition of what constitutes a defective item or an error depends on the nature of the test performed by the inspector and on the goals of the inspection. Clearly, the inspector will want to avoid placing too much emphasis on trivial errors or discrepancies of little practical significance. On the other hand, the value of the inspection is diminished if the inspector is overly permissive in accepting errors. A proper balance must be maintained between these extremes, and considerable judgment is required. When the attribute test is based on a measurement, a defective is usually defined as a discrepancy too large to be explained by the normal variability of the test measurement. The error effects associated with the measurement system are not defects in the attribute sense. However,

defects can be defined by a systematic procedure for determining that an observed difference is unlikely to have been caused by measurement variability alone.

The salient features of attribute sampling may be summarized as follows:

1. A test (measurement, examination) is used to classify each item as either defective or not.

2. The emphasis is on the rate of occurrence of defects or on the number of defects rather than on the magnitude of the observed discrepancies.

3. The conclusion of an attribute sampling procedure is usually a determination to accept or reject the inspected population as satisfactory in terms of quality.

4. Poor quality (high defect rate or number) may indicate either a breakdown of the internal control system or a possible loss or diversion of material. Follow-up investigations are needed to determine the cause of the observed discrepancies.

5. Since attribute sampling involves counted values (number of defects, number of items), discrete probability distributions are used in modeling the procedure.

6. The decision to accept or reject can be treated as a statistical hypothesis test. Alternatively, the number or proportion of defects in the population can be estimated.

Two discrete distributions are in wide use for attribute sampling applications, the hypergeometric distribution and the binomial distribution. The Poisson distribution has also seen some use, primarily as an approximation to the binomial distribution. With the advent of hand-held calculators and readily available computing facilities, the need for the Poisson approximation has diminished. These distributions are introduced and illustrated in Chapter 2. Hypothesis tests based on the binomial distribution are presented in Chapter 4.

17.2.2 Applications of the Hypergeometric Distribution

As discussed in Chapter 2, the hypergeometric distribution is an appropriate model for random sampling *without replacement* from a finite population, and this is the justification for using it in verification applications. Although the hypergeometric model is exact in this context, it is somewhat cumbersome to use when computing specific probabilities. The use of the binomial distribution as a more tractable but approximate model will be discussed in the next section.

The hypergeometric distribution can be applied to obtain exact solutions to the following kinds of questions:

1. **Calculating probabilities of detection.** Consider a population of known size and with a known fraction of defective items in the population.

For random samples of various sizes without replacement, what are the probabilities of obtaining various numbers of defective items in the sample? In other words, given a specified sampling procedure, what is the probability of observing various numbers of defective items?

2. **Designing a sampling plan.** Consider a population of known size. The actual fraction of defects in the population is not known. The inspector wishes to design a sampling plan so that (1) if the actual number or fraction of defects is greater than or equal to some specified value, the population will be rejected with some specified (usually high) probability, and (2) if the actual number or fraction of defects is less than or equal to some specified value, the population will be rejected only with a specified low probability. How large should the sample size be to satisfy these conditions?

Example 17.2 To illustrate the computations involved, consider the simple (and unrealistic) example where the population consists of 20 cans of UO_2 powder. The fraction defective in the population is assumed to be 10%, or two cans. For a sample of size 5, calculate the probability of obtaining various numbers of defectives in the sample.

First, notation is defined and the facts given in the problem statement are recorded.

N \approx population size = 20

n \approx sample size = 5

p \approx fraction defective in the population = 0.1

D \approx number of defectives in the population = $Np = (20)(0.1) = 2$

X \approx number of defectives in the sample (random variable)

x \approx number of defectives in the sample (observed value of random variable) = 0, 1, or 2.

From Equation 2.17, the formula for calculating the required probability is

$$\Pr(X = x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad (17.1)$$

where the notation $\binom{a}{b}$, introduced in Chapter 2, indicates

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} \quad (17.2)$$

For this example, the chance of getting zero defectives in the sample is

$$\Pr(X = 0) = \frac{\binom{2}{0} \binom{18}{5}}{\binom{20}{5}} = \frac{\frac{2!}{0!2!} \frac{18!}{5!13!}}{\frac{20!}{5!15!}}$$

By cancellation of identical factors, this expression simplifies to†

$$\Pr(X = 0) = \frac{15(14)}{20(19)} = 0.5526$$

Repeating the calculation for $x = 1$ and $x = 2$,

$$\begin{aligned} \Pr(X = 1) &= \frac{\binom{2}{1} \binom{18}{4}}{\binom{20}{5}} = \frac{\frac{2!}{1!1!} \frac{18!}{4!14!}}{\frac{20!}{5!15!}} \\ &= \frac{2(5)(15)}{20(19)} = 0.3947 \end{aligned}$$

$$\begin{aligned} \Pr(X = 2) &= \frac{\binom{2}{2} \binom{18}{3}}{\binom{20}{5}} = \frac{\frac{2!}{2!0!} \frac{18!}{3!15!}}{\frac{20!}{5!15!}} \\ &= \frac{5(4)}{20(19)} = 0.0526 \end{aligned}$$

Thus, under the specified conditions, there is a 55% chance that the sample will contain no defective items, and a 45% chance that it will contain one or more.

For realistic population sizes and sample sizes, hand calculation of hypergeometric probabilities quickly becomes unwieldy. Modern computers and hand calculators both evaluate factorials easily, and most software packages contain programs for computing the terms of a hypergeometric

†Recall from Chapter 2 that $0! = 1$.

distribution. Logarithms can also be used in evaluating factorial terms, since with logarithms all of the numbers remain within reasonable bounds.

Continuing with this simple example, suppose that the fraction defective in the population is unknown. The inspector uses the criterion that a population fraction defective of 100 $p_A = 20\%$ or higher is unacceptable and should be rejected with high probability (say 0.9) in the course of inspection. In addition, the inspector uses the criterion that a population fraction defective of 100 $p_0 = 5\%$ or lower is acceptable and should be accepted with high probability (say 0.9) in the course of inspection. How large should the sample size be to satisfy these criteria?

The statistical formulation of this problem will be in terms of hypothesis testing. Specifically, the question to be answered is: What sample size should be selected to induce the desired power in the statistical test? The reader may wish to refer to Chapter 4 to review the basic concepts of hypothesis testing.

The null hypothesis is that the population fraction defective is less than or equal to 5%. The alternative hypothesis is that the population fraction defective is greater than or equal to 20%. The probability of rejecting the null hypothesis when it is true is to be 0.1 in this example, because the probability of accepting the null hypothesis when it is true was chosen as 0.9. The probability of rejecting the null hypothesis when the alternative holds was chosen to be 0.9. This is the power of the test, or equivalently, one minus the probability of a Type II error.

The hypothesis test requires the selection of a critical value, U , which represents the largest number of defectives that can appear in the sample without rejecting the null hypothesis and the population. U is often called the "acceptance number." If the sample contains U or fewer defectives, the population is accepted. By definition, the Type I error probability is

$$\Pr(X > U | p = p_0) = \alpha \quad (17.3)$$

and the Type II error probability is

$$\Pr(X \leq U | p = p_A) = \beta \quad (17.4)$$

Using the hypergeometric distribution to determine these probabilities, the equations become

$$\Pr(X > U | p = p_0) = \sum_{x=U+1}^{D_0} \frac{\binom{D_0}{x} \binom{N-D_0}{n-x}}{\binom{N}{n}} = \alpha \quad (17.5)$$

$$\Pr(X \leq U | p = p_A) = \sum_{x=0}^U \frac{\binom{D_A}{x} \binom{N-D_A}{n-x}}{\binom{N}{n}} = \beta \quad (17.6)$$

where

N = population size = 20

n = sample size (to be determined)

U = acceptance number (to be determined)

D_0 = number of defectives in population under the null hypothesis = $Np_0 = 20(0.05) = 1$

D_A = number of defectives in population under the alternative hypothesis = $Np_A = 20(0.20) = 4$.

The design problem is to determine the minimum value of n , and the corresponding value of U , for which certain predetermined error probabilities α_0 and β_0 are attained. Clearly, setting $n = N$ and $U = 1, 2, \text{ or } 3$ will make both α and β zero and satisfy the conditions. However, the intent is to satisfy the conditions with the smallest possible sample. Because n and U must be integers, the required solution may not produce values of α and β which are exactly equal to the desired error probabilities α_0 and β_0 . Thus, the problem is to determine the minimum integer value of n , and the corresponding U , for which $\alpha \leq \alpha_0$ and $\beta \leq \beta_0$. Tables (Sherr, 1972) are available to assist in solving this problem, and many software packages contain programs for its solution. Various approximations are also available (Jaech, 1973, Section 9.1).

In the present simple example, a solution can be determined by choosing successively larger values of U beginning with $U = 0$, determining the values of n for which the value of α computed from Equation 17.5 is less than or equal to $\alpha_0 = 0.1$, and then determining the minimum n in this set for which the value of β computed from Equation 17.6 is less than or equal to $\beta_0 = 0.1$. For $U = 0, D_0 = 1$,

$$\alpha = \frac{\binom{19}{n-1}}{\binom{20}{n}} = \frac{n}{20}$$

so that $\alpha \leq 0.1$ if $n = 1$ or 2 . For $U = 0, D_A = 4$,

$$\beta = \frac{\binom{16}{n}}{\binom{20}{n}} = \frac{(20-n)(19-n)(18-n)(17-n)}{(20)(19)(18)(17)}$$

For $n = 1$, $\beta = 0.80$; for $n = 2$, $\beta = 0.632$. Thus, for $U = 0$ there is no value of n for which $\alpha \leq 0.1$ and $\beta \leq 0.1$. If $U = 1$, then for $D_0 = 1$, $\alpha = 0$, because the occurrence of two or more rejects is not possible. Thus, any value of n will satisfy the requirement $\alpha \leq 0.1$. For $U = 1$, $D_A = 4$,

$$\beta = \frac{\binom{4}{0} \binom{16}{n}}{\binom{20}{n}} + \frac{\binom{4}{1} \binom{16}{n-1}}{\binom{20}{n}}$$

$$= \frac{(20-n)(19-n)(18-n)(17+3n)}{(20)(19)(18)(17)}$$

For $n = 13$, $\beta = 0.101$; for $n = 14$, $\beta = 0.061$, so that the smallest sample size for which the desired conditions are possible is $n = 14$, with an acceptance number of 1. The actual error levels are $\alpha = 0$, $\beta = 0.06$, rather than the desired value of $\alpha_0 = \beta_0 = 0.1$. This behavior is typical of small populations and small numbers of defects. When both the size of the population and the number of defects present are large, solutions closely approximating the required conditions can be obtained.

Note that for $U = 2$, any sample size will again satisfy the conditions on α . However, the expression for β will contain an additional positive term, so that for a given n , β will always be larger than its value for $U = 1$. Hence, the minimum n for which $\beta \leq \beta_0$ must be greater than or equal to that for $U = 1$.

In safeguards applications of attribute sampling, the special case $D_0 = 0$, which corresponds to the absence of any defective items in the population being inspected, is frequently an appropriate null hypothesis. In this case, the occurrence of any defect in a sample, regardless of the sample or population size, would be unacceptable, so that $U = 0$ is the only possible acceptance number. When this is the case, the computations needed to calculate the appropriate sample size are greatly simplified. First, note that the probability α of a Type I error is zero; it is impossible to find any defectives if the population does not contain any, so the chance of rejection when the null hypothesis is true is zero. Second, the Type II error probability β depends on only one term of the summation shown earlier (Equation 17.6), namely the term for which $x = 0$. Then

$$\begin{aligned}
 \beta &= \frac{\binom{D_A}{0} \binom{N-D_A}{n-0}}{\binom{N}{n}} = \frac{D_A! \frac{(N-D_A)!}{0! D_A! n! (N-D_A-n)!}}{\frac{N!}{n!(N-n)!}} \\
 &= \frac{(N-D_A)!(N-n)!}{N! (N-D_A-n)!} \\
 &= \frac{(N-n)(N-n-1)(N-n-2) \dots (N-D_A-n+1)}{N (N-1) (N-2) \dots (N-D_A+1)} \quad (17.7)
 \end{aligned}$$

Note that for $N = 20$, $n = 1$, and $D_A = 4$, Equation 17.7 reduces to the result given for β when $U = 0$ in Example 17.2. The value $U = 0$ may occur as a solution to the design problem even when D_0 is not zero. Equation 17.7 is also the basis for the approximation (Jaech, 1973, p. 327)

$$\beta \cong \left[1 - \frac{2n}{2N - D_A + 1} \right]^{D_A}$$

The corresponding approximation for the sample size for a given value of β is

$$n = 0.5(1 - \beta^{1/D_A})(2N - D_A + 1) \quad (17.8)$$

The use of a zero acceptance number has considerable merit in audit and inspection applications. In many cases, the emphasis may properly be placed on uncovering errors, if they exist, rather than on attempting to discriminate between the acceptable and rejectable quality levels. In financial auditing, sampling plans of this type are called "discovery sampling plans," which is suggestive of their emphasis on finding errors rather than testing a hypothesis.

Example 17.3 Returning to Example 17.2 and applying the zero acceptance number formula with

$$\beta = 0.1$$

$$D_A = 4$$

$$N = 20$$

yields

$$\begin{aligned} n &= 0.5[1 - (0.1)^{0.25}](40 - 4 + 1) \\ &= 8.1 \end{aligned}$$

Fractional sample sizes should always be rounded up, so for this example a sample of size 9 is needed to ensure a 0.9 probability of having at least one defective in the sample, and thus rejecting the population when 4 of 20 items in the population are defective. (The actual values are $\beta = 0.102$ for $n = 8$ and $\beta = 0.068$ for $n = 9$.)

Example 17.4 Stratum 3 of the material balance considered in Example 14.6 consisted of 2770 containers of solid waste. After the content is determined these containers are sealed and stored, frequently for some period of time, until they are discarded or processed. Inspectors checking the integrity of the seals wish to examine a sufficient sample of the containers to infer, based on the absence of any defective seals in the sample ($U = 0$), that the probability is 0.95 that 99% of the seals are intact. For $N = 2770$, $\beta = 0.05$, and $D_A = (2770)(0.01) = 27.7$, Equation 17.8 yields the result

$$\begin{aligned} n &= 0.5[1 - (0.05)^{1/27.7}](5540 - 27.7 + 1) \\ &= 282.6 \end{aligned}$$

so that a sample of 283 containers selected at random should be inspected. Note that rounding D_A up to 28 would lead to a sample size of 280, which is nonconservative (see Example 17.7).

If several distinct populations are to be sampled in an inspection, e.g., several different strata of material, some attention should be given to the problem of multiple testing and simultaneous inference (see Chapter 10). In the zero acceptance number case, the situation is simple because the false alarm (Type I error) probability is zero (under the null hypothesis) for every stratum, so that the overall probability is also zero. More generally, however, the situation with multiple strata is not so simple and control of the overall Type I error probability requires more detailed analysis.

17.2.3 Applications of the Binomial Distribution

The binomial distribution is more widely used in attribute sampling applications than is the hypergeometric. The hypergeometric distribution

provides a theoretically exact model for sampling without replacement from a finite population. However, the binomial distribution provides an excellent approximation in many situations and is generally easier to work with.

The binomial distribution is theoretically exact when sampling is with replacement or when sampling is from an infinite population. Sampling with replacement is not realistic in safeguards applications—to verify the same item more than once would be wasteful (unless it were done as a quality control procedure, which is another matter)—but, in practice, populations to be verified are frequently large enough to be effectively infinite. An intuitive explanation is given in Section 2.4.2.

In safeguards, the most common application is with zero acceptance number sampling. The question of interest is to determine the probability of obtaining one or more defective items in the sample, given a specified fraction defective in the population. The probability of obtaining one or more defective items is one minus the probability of obtaining no defective items. Applying the binomial distribution (see Equation 2.20),

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (17.9)$$

with

- x = number of defectives in sample
- p = fraction defective in the population
- n = sample size

$$\begin{aligned} \Pr(X = 0) &= \binom{n}{0} p^0 (1-p)^n \\ &= (1-p)^n \end{aligned} \quad (17.10)$$

This formula can be interpreted as the probability of no successes (defectives in the sample) in n independent trials (items sampled) where the chance of success at each trial is p , the probability that a randomly selected item is defective.

A preferred, alternative form of the binomial approximation involves a switch in viewpoint. Instead of looking at n independent drawings of items from a population consisting of $D_A = Np$ defectives and $N - D_A$ non-defectives, consider the assignment of D_A defectives to a population consisting of $n = Nf$ sampled items and $N - n$ non-sampled items. Then the probability of x successful assignments of a defective to a sampled item, which has a probability f of occurrence, is

$$\Pr(X = x) = \binom{D_A}{x} f^x (1-f)^{D_A-x}$$

When the acceptance number is zero,

$$\Pr(X = 0) = (1-f)^{D_A} \quad (17.11)$$

This can be interpreted as the probability that each of the D_A defects escapes the sample.

It is useful to compare these two binomial approximations to develop insight into how well they work and the conditions under which they break down.

Example 17.5 In Example 17.2 of the preceding section, the target population consisted of 20 cans of UO_2 powder, and the assumed fraction defective was 10%, so that $D_A = 2$. The sample size was $n = 5$ and the fraction sampled was $f = 0.25$. The probability of obtaining zero defectives in the sample according to the (exact) hypergeometric distribution was

$$\Pr(X = 0) = 0.5526$$

By the two binomial approximations,

$$\Pr(X = 0) = (1-p)^n = (1-0.1)^5 = 0.5905$$

$$\Pr(X = 0) = (1-f)^{D_A} = (1-0.25)^2 = 0.5625$$

The three results agree more closely than one might have expected, given that the fraction sampled is as large as 0.25. Note also that the second binomial approximation is noticeably better than the first, and that both approximations overestimate the probability that no defectives will be found. Equivalently, they underestimate the probability of detection. In this sense, they are conservative approximations.

Example 17.6 Suppose a stratum contains 200 plutonium fuel pins, 10 of which are defective. The results obtained with the binomial approximations are to be compared with those obtained using the exact hypergeometric model. Let

$$\begin{aligned}
 n &= \text{sample size} \\
 N &= \text{population size} = 200 \\
 f &= \text{fraction sampled} = n/N = n/200 \\
 D_A &= \text{number of defectives} = 10 \\
 p &= \text{fraction defective} = D_A/N = 0.05
 \end{aligned}$$

Then, from Equations 17.10, 17.11, and 17.7, the results in Table 17.2 can be calculated. The binomial approximations break down when the fraction sampled is large, although $(1-f)^{D_A}$ seems to work better than $(1-p)^n$.

TABLE 17.2
Comparison of Approximations for Example 17.6

| n | f | $(1-p)^n$ | $(1-f)^{D_A}$ | $\frac{\binom{N-D_A}{n}}{\binom{N}{n}}$ |
|-----|-------|-----------|---------------|---|
| 25 | 0.125 | 0.2774 | 0.2631 | 0.2545 |
| 50 | 0.25 | 0.0769 | 0.0563 | 0.0521 |
| 100 | 0.5 | 0.0059 | 0.0010 | 0.000771 |
| 150 | 0.75 | 0.000456 | 0.000001 | 0.00000046 |

Perhaps the most prevalent application of the binomial approximation is in calculating sample sizes. Starting from Equation 17.11

$$\Pr(X = 0) = (1-f)^{D_A}$$

and substituting $\Pr(X = 0) = \beta$ and $f = n/N$, the following formulas can be obtained:

$$\beta = (1-f)^{D_A} \quad (17.12)$$

$$f = 1 - \beta^{1/D_A} \quad (17.13)$$

$$n = N(1 - \beta^{1/D_A}) \quad (17.14)$$

These formulas are easy to apply with a hand calculator, and form the basis for IAEA procedures for determining sample size (IAEA, 1988).

Because the binomial approximation underestimates the probability of detection, it will overestimate the sample size required to achieve a given probability of detection. Thus, Equation 17.14 is a conservative estimate of the necessary sample size in the sense that it will always be larger than needed to achieve the required detection probability or, equivalently, will actually produce a detection probability higher than that required. The estimate may be unduly conservative if the defective items constitute more than about 10% of the items in the population, and Equation 17.8 should be used to compute the sample size (see Example 17.8).

Example 17.7 In a highly enriched uranium fuel fabrication plant, one stratum in the physical inventory contains 53 cylinders of uranium hexafluoride (UF_6). The stratum contains a total of 384.8 kg ^{235}U . Thus, the average item contains 7.26 kg ^{235}U . If the goal quantity were 5 kg ^{235}U , then a gross discrepancy in a single cylinder would constitute an unacceptable loss or diversion. In order to have a 0.95 probability of detecting a single defective, clearly 95% of the items would have to be inspected. The formula correctly gives that result because

$$\begin{aligned} n &= N(1 - \beta^{1/D_A}) \\ &= 53(1 - 0.05^1) \\ &= 50.35 \end{aligned}$$

Because fractional sample sizes are always rounded up, the sample size is rounded to 51. In practice, when such a large fraction of the population is to be verified, it is often logistically and administratively simpler to sample 100% of the population, ignoring the statistical sampling plan.

Suppose now that the goal quantity is 25 kg ^{235}U , as might be the case in an IAEA inspection. This translates to

$$D_A = 25 \text{ kg} / 7.26 \text{ kg/item} = 3.44$$

defective items in the population. Because by definition an item either is or is not defective, it is tempting to round D_A to an integer value. However, to round up to $D_A = 4$ is nonconservative—it may lead to a detection probability smaller than the specified target value. On the other hand, rounding down may unduly inflate the required sample size, especially when (as in this case) the number of defectives is small. For design purposes, it is desirable simply to use the approximate formulas in Equations

17.8, 17.12, 17.13, and 17.14 with a fractional value of D_A . This recognizes that it is actually the proportion, not the integral number, of defectives which is of concern in most safeguards situations. The sample size calculation is

$$\begin{aligned} n &= N(1 - \beta^{1/D_A}) \\ &= 53[1 - (0.05)^{0.29}] \\ &= 30.8 \end{aligned}$$

This is then rounded up to $n = 31$.

Example 17.8 In a low-enriched uranium fuel fabrication plant, the inventory at the end of a material balance period includes 1000 trays of pellets. Each tray contains 5 kg uranium enriched to 3% ^{235}U . This translates to 150 g of ^{235}U per tray. Assuming the goal quantity is 75 kg ^{235}U , 500 of the 1000 trays would have to be gross defects (i.e., contain no ^{235}U at all) to constitute a total discrepancy for the stratum of one goal quantity. To detect at least 1 of the 500 defective items with probability 0.95, the sample size should be

$$\begin{aligned} n &= N(1 - \beta^{1/D_A}) \\ &= 1000[1 - (0.05)^{0.002}] \\ &= 5.97 \end{aligned}$$

which is rounded up to obtain $n = 6$.

Because the fraction defective is quite high (50%), it is interesting to check the adequacy of the binomial approximation with a sample size of 6. From Equation 17.12, the probability of not detecting the discrepancy is

$$\beta = (1-f)^{D_A} = \left[1 - \frac{6}{1000}\right]^{500} = 0.0494$$

while Equation 17.7 results in the value

$$\beta = \frac{\binom{N-D_A}{n}}{\binom{N}{n}} = \frac{\binom{500}{6}}{\binom{1000}{6}} = 0.0154$$

Once again, it can be seen that the approximate formula is conservative and gives noticeably different results than the exact solution when the fraction defective is large. In this case, the use of Equation 17.8 would significantly improve the sample size calculation. The result is

$$\begin{aligned} n &= 0.5 [1 - (0.05)^{0.002}] (2000 - 500 + 1) \\ &= 4.48 \end{aligned}$$

which rounds to a sample size of 5. From Equation 17.7, the exact value of β for a sample size of 5 is

$$\beta = \frac{\binom{500}{5}}{\binom{1000}{5}} = 0.0309$$

17.3 INFERENCES FROM MEASUREMENT DIFFERENCES

This section considers the analysis of the actual measured differences between declared values of sampled items and remeasurements by the inspecting agency. As with attributes, both *compliance* tests to determine the status of the measurement systems and *substantive* tests of the presence of the material can be based on this type of data. The statistical techniques are primarily extensions of the techniques for modeling and estimating measurement errors considered in Chapters 14 and 15.

17.3.1 Inferences About Measurement Errors

The role of the measurement system in nuclear materials accounting has been described in detail in Chapter 13. It was pointed out there that the utility of the entire nuclear material accounting system is heavily dependent on the capability and performance of the measurement system. For this reason, nuclear material accounting systems normally include a measurement control program, as outlined in Section 13.3.2.1, whose basic function is to monitor and control the precision and accuracy of measurement. Statistical methods applicable to the routine operations of a measurement control program have been presented in Chapter 15.

Measurement control programs normally make provisions for some kind of audit or review activity to confirm the capability and performance of the measurement system. Such audits may include both tests of compliance with established procedures and direct tests of measurement perfor-

mance. The applicable statistical techniques are essentially the same as those described in Chapter 15 for monitoring measurement systems. Attribute tests can also be used for compliance testing purposes. The attribute of interest is usually the adherence to prescribed measurement procedures. In any audit of this kind, the distinguishing characteristic is the provision of an independent check of routine operations.

While reliance on a strong measurement control program may provide the primary assurance concerning measurement capability and performance, other sources of information also exist by which measurement capability and performance can be verified. In particular, paired data (either shipper–receiver or operator–inspector data) can be used for this purpose. The following sections describe how inspection data in the form of operator–inspector paired comparisons can be used to assess the operator’s measurement performance.

17.3.1.1 Use of Inspection Measurements to Verify Measurement Capability and Performance

One of the many activities carried out in the course of an inspection consists of independent measurements performed by the inspection team on a sample of items. (Recall from Chapter 13 the definition of an item as the primary accounting unit in the material accounting system—an object or container that has a unique material quantity assigned to it and that can be identified and accounted for as a separate entity.) Such inspection measurements are multipurpose, as can be seen in the other sections of this chapter. One of the purposes, and the only one discussed in this section, is to provide information that can be used to test the capability and performance of the operator’s measurement system.

The basic data are in the form of paired measurements on individual items. The operator will have assigned a value (material quantity) to every item entering into the material balance. The inspector selects a sample of items and measures them. For each item sampled, the operator’s and inspector’s values may be compared, and the difference between the two values computed. The observed differences will reflect the combined effect of the operator’s and inspector’s measurement errors. Highly precise and unbiased measurements should give rise to only small differences. Highly precise but biased measurements will tend to yield consistent but nonzero differences. Imprecise measurements should result in highly variable differences. Thus, the observed differences contain information that can be used to formalize the assessment process.

In practice, the sequence of inspection activities is the following. On the basis of past experience, the inspection team sets design criteria for this phase of the inspection program. Using the design criteria, an inspection plan is drawn up, including a sampling plan for selecting items for verification. Next, the inspection plan is implemented by selecting items at

random and making independent measurements on the sampled items. The measurement results are next compared to the operator's reported values and the performance of the operator's measurements assessed. Finally, depending on the outcome of the assessment, a follow-up investigation of some kind may be initiated. In any case, conclusions and recommendations concerning this phase of the inspection are documented and transmitted through the appropriate channels. In Section 17.3.1.2, the procedures for evaluating the inspection data are presented. In Section 17.3.1.3, the procedures for determining an appropriate sample size are treated.

17.3.1.2 Evaluating Inspection Data to Assess Measurement Performance

The analysis described here characterizes the variability of the differences between the operator's and the inspector's values for some sample of items from a single stratum. The variability actually observed will be compared to the variability that would be expected if actual measurement performance were in fact consistent with stated or prescribed measurement capability.

Assume that the inspector has sampled n items of the N items in the stratum, and that measured values for each item have been obtained and recorded for both operator and inspector. Let

$$d_k = x_{1k} - x_{2k} = \text{observed difference for item } k \quad (17.15)$$

where

x_{1k} = operator's measurement result for item k

x_{2k} = inspector's measurement result for item k

$$\bar{d} = \frac{1}{n} \sum_{k=1}^n d_k = \text{observed mean difference}$$

$$s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (d_k - \bar{d})^2 = \text{estimated variance of the differences}$$

The assumed error models for operator and inspector results are

$$x_{1k} = \mu_k + \beta_1 + \epsilon_{1k}$$

$$x_{2k} = \mu_k + \beta_2 + \epsilon_{2k} \quad (17.16)$$

where μ_k is the true amount of material in item k , β_1 and β_2 are the constant error effects associated with the operator and inspector, respectively,

and ϵ_{1k} and ϵ_{2k} are the operator's and inspector's individual error effects for item k . For this treatment, only two error sources are considered for each measurement, one affecting all items and the other each item. If this assumption is not valid, then the differences d_k must be modeled and analyzed using the more complex hierarchical analysis of variance models introduced in Section 5.3.1 and applied in Section 14.3. It follows from Equations 17.15 and 17.16 that

$$d_k = \beta_1 - \beta_2 + \epsilon_{1k} - \epsilon_{2k}$$

and assuming $E(\epsilon_{ik}) = E(\epsilon_{2k}) = 0$, $\text{Var}(\epsilon_{1k}) = \sigma_1^2$, $\text{Var}(\epsilon_{2k}) = \sigma_2^2$, and $\text{Cov}(\epsilon_{1k}, \epsilon_{2k}) = \sigma_{12}$, it follows that

$$E(d_k) = E(\bar{d}) = \beta_1 - \beta_2$$

and

$$E(s_d^2) = \sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

If the error effects are also assumed to have a normal distribution, the procedure provided in Section 4.2.3 can be applied to test the hypothesis that the variance σ_d^2 of the differences is equal to a specified value, corresponding to stated or prescribed levels of measurement capability.

The procedures of Section 4.4 can be applied to test the hypothesis $\beta_1 - \beta_2 = 0$ that there is no relative bias between the inspector and the operator in this sample. This is a test of the absence of a constant effect in the operator's measurements only if the inspector's measurements are known or assumed to be unbiased. In this case, the inspector's measurement is being considered a standard in the sense discussed in the introduction to Section 15.2. Numerically this test is equivalent to an F-test, with 1 and $n-1$ degrees of freedom, of the hypothesis that the variance component associated with the constant effects β_1 and β_2 is zero. This formulation may be more appropriate when considering these effects as a sample from a large population. When the assumed error model is more complex, additional levels of fixed or random effects must be included in the analysis.

Example 17.9 During an inspection at a low-enriched uranium fuel fabrication plant, 19 cans of UO_2 powder are randomly selected for verification by the inspector. After the items are measured by the inspector, the

observed differences (in grams of uranium) are calculated. The data are listed in Table 17.3.

TABLE 17.3
Operator-Inspector Differences
for a Sample of
UO₂ Powder Cans

| Item
k | Difference
(g uranium) |
|-----------|---------------------------|
| 1 | -32 |
| 2 | -10 |
| 3 | -72 |
| 4 | 8 |
| 5 | 40 |
| 6 | 32 |
| 7 | -71 |
| 8 | 40 |
| 9 | 33 |
| 10 | 25 |
| 11 | -31 |
| 12 | -20 |
| 13 | 42 |
| 14 | 12 |
| 15 | 13 |
| 16 | -4 |
| 17 | -1 |
| 18 | 0 |
| 19 | -22 |

The plant operator, on the basis of an extensive measurement control program, claims a measurement error variance for this type of measurement of $\sigma_1^2 = 361 \text{ (g)}^2$. The inspector, from his own measurement control program, estimates his independent measurement error variance to be $\sigma_2^2 = 400 \text{ (g)}^2$. The question is whether the variability of the differences in Table 17.3 is consistent with the stated random error variances.

One approach is to perform a chi-square test of the variance. This test was introduced in Section 4.2.3. The null hypothesis is that the true variance of the differences is (because $\sigma_{12} = 0$) $\sigma_d^2 = 361 \text{ (g)}^2 + 400 \text{ (g)}^2 = 761 \text{ (g)}^2$. The alternative is that the variance is either larger or smaller than the stated value. To test the hypothesis, the test statistic

$$\chi^2 = \frac{(n-1) s_d^2}{\sigma_d^2} \quad (17.17)$$

is used (Equation 4.21). Because

$$\begin{aligned}(n-1)s_d^2 &= \sum(d_k - \bar{d})^2 = \sum d_k^2 - \frac{1}{n} (\sum d_k)^2 \\ &= 21,272.95 (g)^2\end{aligned}$$

the test statistic takes on the value

$$\begin{aligned}\chi^2 &= 21,272.95/761 \\ &= 27.95\end{aligned}$$

This value is compared to upper and lower $100\alpha/2$ percentage points of the chi-square distribution with 18 degrees of freedom. From Table A4 for $\alpha = 0.05$,

$$\chi_{0.025}^2(18) = 8.23$$

$$\chi_{0.975}^2(18) = 31.5$$

Because the observed value of the test statistic falls within this range, the null hypothesis is not rejected, and it is concluded that the observed variability of the differences in Table 17.3 is not inconsistent with the stated measurement capabilities.

As pointed out in Chapter 4, the chi-square test is sensitive to departures from the assumed normal distribution of the differences. This is an example of the general principle succinctly stated by Kendall and Stuart (1979, p. 493): "Tests on means are robust; by comparison, tests on variances can only be described as frail." The normal theory test procedure, when applied to nonnormal data, may result in a Type I error probability different from the specified $\alpha = 0.05$. In practice, this means that incorrect conclusions may be drawn more often than is acceptable.

The stem-and-leaf diagram for the differences in Table 17.3 shown in Table 17.4 suggests that the differences -71 and -72 might be outliers, and also raises some question as to the distribution of the remaining differences. Powder cans of this type typically contain about 20 kg of uranium and can usually be measured with a relative standard deviation of less than 0.1%, which is consistent with the claims of both inspector and operator. If the two extreme values are excluded, the estimated value of σ_d^2 based on 16 degrees of freedom is $s_d^2 = 634.11 (g)^2$ and the mean

TABLE 17.4
Stem-and-Leaf Display of
Operator-Inspector Differences

| | |
|----|-----|
| -7 | 12 |
| -6 | |
| -5 | |
| -4 | |
| -3 | 12 |
| -2 | 02 |
| -1 | 0 |
| -0 | 14 |
| 0 | 08 |
| 1 | 23 |
| 2 | 5 |
| 3 | 23 |
| 4 | 002 |
| 5 | |

difference is 7.35 g. For this estimate of the variance, the value of χ^2 computed from Equation 17.17 is

$$\begin{aligned}\chi^2 &= \frac{(16)(634.11)}{761} \\ &= 13.33\end{aligned}$$

which is not only not significant, but is also slightly smaller than its expected value of 16. The mean difference of 7.35 is also not significant, because

$$t = \frac{7.35\sqrt{16}}{\sqrt{634.11}} = 1.168$$

is well within the 5% critical value of Student's *t*-distribution with 16 degrees of freedom. In spite of the lack of significance of the overall chi-square test at the 0.05 level, it seems reasonable to conclude that the values -71 and -72 are unusual and should be investigated, while the remaining differences, in spite of the lack of a characteristic bell-shaped distribution, represent normal measurement performance.

17.3.1.3 Determining the Sample Size

The procedure for determining the sample size needed to attain a specified power level (or equivalently, Type II error probability) is given in Section 4.3.2. In particular, see Example 4.9. To calculate the sample size it is necessary to specify:

1. The magnitude of departure from the assumed measurement capability to be detected.
2. The probability with which such a departure, if it exists, is to be detected (i.e., the power of the test).
3. The allowable probability of incorrectly concluding that a departure exists when in fact it does not (i.e., the Type I error or false alarm probability).

Example 17.9 (cont'd)

To continue Example 17.9, suppose that the inspection team has determined that there would be cause for concern if the operator's measurement performance, as measured by his measurement error standard deviation, was a factor of two larger than prescribed capability levels, where the capability levels might represent widely accepted standards for the variability of a particular combination of measurement techniques. The inspection team decides that if such a deviation occurs, the probability of detecting it should be 0.9 and an allowable false alarm probability for such a test is 0.05.

If the operator's standard deviation is higher than the prescribed level by a factor of two, then his variance will be high by a factor of $2^2 = 4$. Under the null hypothesis that performance equals capability, the variance of the difference is

$$\sigma_d^2 = \sigma_1^2 + \sigma_2^2$$

Under the alternative hypothesis,

$$\sigma_{da}^2 = 4\sigma_1^2 + \sigma_2^2$$

Using the values assumed for this example,

$$\sigma_1^2 = 361 \text{ (g)}^2$$

$$\sigma_2^2 = 400 \text{ (g)}^2$$

$$\sigma_d^2 = 761 \text{ (g)}^2$$

$$\sigma_{da}^2 = 4(361) + 400 = 1844 \text{ (g)}^2$$

The ratio of σ_{da}^2 to σ_d^2 , denoted by λ in Chapter 4, is

$$\lambda = \frac{\sigma_{da}^2}{\sigma_d^2} = \frac{1844}{761} = 2.42$$

Entering Table A4 with $\alpha = 0.05$, $\beta = 0.1$, and $\lambda = 2.42$, and seeking the row in which

$$\lambda \chi_{\beta}^2 \cong \chi_{1-\alpha}^2$$

or, in this example

$$2.42 \chi_{0.1}^2 \cong \chi_{0.95}^2$$

the following table can be constructed:

TABLE 17.5
Selected Chi-Square Values

| Degrees of freedom | $\chi_{0.95}^2$ | $\chi_{0.1}^2$ | $\chi_{0.95}^2 / \chi_{0.1}^2$ |
|--------------------|-----------------|----------------|--------------------------------|
| 25 | 37.6525 | 16.4734 | 2.29 |
| 23 | 35.1725 | 14.8479 | 2.37 |
| 21 | 32.6705 | 13.2396 | 2.47 |
| 22 | 33.9244 | 14.0415 | 2.42 |

Thus, the desired power will be attained with 22 degrees of freedom, which implies a sample size one larger, or a requirement for 23 observed differences.

In calculating sample sizes, it was assumed that the parameters α , β , and λ (the ratio of the variance under the alternative hypothesis to that under the null hypothesis) were given. The parameters α and β present little difficulty, as they will usually be assigned one of the conventional values 0.1, 0.05, or 0.01. The parameter λ is less straightforward. It is not easy to determine how large a departure from prescribed measurement capability should be cause for concern and should be detected with high

probability. In practice, there is likely to be an economic constraint to consider; total inspection resources are usually limited, and the amount of effort devoted to tests of measurement capability must be balanced against the effort needed for other inspection activities. From a statistical viewpoint, the most useful approach is to present a power curve or operating characteristic curve as described in Chapter 4, so that the nature of the tradeoff can be readily assessed. The final decision about how much effort to devote to sampling for assessing measurement capability can then be made in light of the available resources.

17.3.2 Inferences About Material Quantities

In this section, techniques are presented that can be used to make direct inferences about material quantities. Whereas the emphasis in the previous section was on *compliance* tests of the internal control system to determine the state of control over errors, the emphasis in this section is on *substantive* tests of the accounts, i.e., direct inferences about inventory differences and amounts of material in particular items, batches, or strata.

The basis for most of the procedures presented in this section is the so-called **difference statistic**. This statistic is computed from paired comparisons (differences) between operator and inspector data; it quantifies, in terms of amounts of material, the agreement or lack of agreement between the quantities recorded by the plant operator and the quantities estimated by the inspector on the basis of independent measurements.

For those readers interested in exploring more of the background and theoretical aspects of these techniques, it is noted that the correction of an inventory difference based on a difference statistic determined from a sample is one of a class of widely used methods in sampling theory that make use of auxiliary information to strengthen the inferences that can be drawn on the basis of a sample. In the context of nuclear material accounting verification, the auxiliary information is supplied by the book values stated by the operator. These book values are used to obtain more powerful tests and more accurate estimates than could be obtained if only the inspector's independent observations were used as the basis for inference. Further details on this subject can be found in the sampling theory literature (Cochran, 1977; Hansen, Hurwitz, and Madow, 1953; and Kish, 1965).

17.3.2.1 Definition of the Difference Statistic

Suppose that during an inspection the inspector independently measures a sample of n_i items from the N_i items in each of the l strata involved in the calculation of the inventory difference (Equation 14.42). From these measurements and the operator's declared values for each item, the differences

$$d_{ik} = x_{1ik} - x_{2ik} \quad (17.18)$$

can be calculated,

where

x_{1ik} = operator's declared values for sampled item k from stratum i

x_{2ik} = inspector's measured value for sampled item k from stratum i

Then the difference statistic \hat{D}_i for stratum i is defined as

$$\hat{D}_i = N_i \bar{d}_i \tag{17.19}$$

where N_i is the total number of items in stratum i , and

$$\bar{d}_i = \frac{\sum_{k=1}^{n_i} d_{ik}}{n_i} \tag{17.20}$$

is the observed mean difference for the sampled items from the i^{th} stratum. The difference statistic for the material balance is formed from the same linear combination of difference statistics that was used to compute the inventory difference from individual stratum totals. From Equation 14.42,

$$\hat{D} = \sum_{i=1}^g c_i \hat{D}_i \tag{17.21}$$

where the coefficients $c_i = \pm 1$ are defined in Section 14.3.5.

If it is assumed that the operator's declared value can differ from the true content of a particular item presented for inspection not only because of measurement error but also due to actual loss or diversion, the $K = 1, \dots, N_i$ declared values for each item in stratum i can be modeled as

$$x_{1iK} = \mu_{iK} + L_{iK} + \epsilon_{1iK} \tag{17.22}$$

where ϵ_{1iK} is the operator's measurement error and L_{iK} is the unknown discrepancy in the K^{th} declared value. For the $k = 1, \dots, n_i$ sampled items, the inspector's measurements contain only an error of measurement, so that

$$x_{2ik} = \mu_{ik} + \epsilon_{2ik} \tag{17.23}$$

If the same conventions are adopted for the error effects due to the inspec-

tor and operator as in Section 14.3.2, then it can be assumed that $E(\epsilon_{1iK}) = E(\epsilon_{2iK}) = 0$ and hence for the sampled items,

$$E(d_{ik}) = \mu_{ik} + L_{ik} - \mu_{ik} = L_{ik} \quad (17.24)$$

Assuming the same notational conventions as in Chapter 14, define the total discrepancy in stratum i as

$$L_i = \sum_{K=1}^{N_i} L_{iK} \quad (17.25)$$

and the total discrepancy in the sample items as †

$$\hat{L}_i = \sum_{k=1}^{n_i} L_{ik} \quad (17.26)$$

From Equations 17.20, 17.24, and 17.26, it follows that

$$\begin{aligned} n_i E(\bar{d}_i) &= \sum_{k=1}^{n_i} E(d_{ik}) \\ &= \sum_{k=1}^{n_i} L_{ik} \\ &= \hat{L}_i \end{aligned} \quad (17.27)$$

The n_i discrepancies L_{ik} in the sampled items are selected randomly from the finite population of true discrepancies of size N_i , so that

$$E'(\hat{L}_i) = n_i L_i / N_i = n_i \bar{L}_i \quad (17.28)$$

where $\bar{L}_i = L_i / N_i$. The prime has been added to indicate that this expectation over the possible samples of n_i items is *not* the same as the expectation E over the possible measurement effects in Equations 17.24 and

† Strictly speaking, as shown in Equation 17.28, it is \hat{L}_i / n_i which is an estimate of L_i / N_i . The "hat" notation is used here for the sums to retain the distinction between sample and population with respect to the total discrepancy present.

17.27. In order that it follow from Equation 17.19 that $E(\hat{D}_i) = L_i$, both Equations 17.27 and 17.28 must hold—i.e., it is necessary that $E(\bar{d}_i) = \hat{L}_i/n_i$ and $E'(\hat{L}_i) = n_i L_i/N_i$. The two expectations are over the measurement and the sampling processes, respectively.

17.3.2.2 The Variance of the Difference Statistic

The estimate $\hat{D}_i = N_i \bar{d}_i$ for a single stratum is a linear combination of the declared values of the operator and the measured values of the inspector. From Equations 17.18 and 17.20, it follows that

$$n_i \bar{d}_i = \sum_{k=1}^{n_i} x_{1ik} - \sum_{k=1}^{n_i} x_{2ik} = x_{1i} - x_{2i} \quad (17.29)$$

or, alternatively, from Equations 17.22, 17.23, and 17.26 that

$$n_i \bar{d}_i = \hat{L}_i + \sum_{k=1}^{n_i} \epsilon_{1ik} - \sum_{k=1}^{n_i} \epsilon_{2ik} \quad (17.30)$$

where x_{1i} and x_{2i} are the totals of the operator's reported values and the inspector's measurements, respectively, and the ϵ_{1ik} are the operator's measurement errors on the sampled items. From Equation 17.29,

$$n_i^2 \text{Var}(\bar{d}_i) = \text{Var}(x_{1i}) + \text{Var}(x_{2i}) - 2 \text{Cov}(x_{1i}, x_{2i}) \quad (17.31)$$

If the operator's and inspector's measurement systems are completely independent, then $\text{Cov}(x_{1i}, x_{2i}) = 0$ and completely separate estimates of $\text{Var}(x_{1i})$ and $\text{Var}(x_{2i})$ can be obtained by identifying the measurement error sources in operator and inspector measurements and applying the methods outlined in Chapter 14 for stratum totals (Section 14.3.2, Equation 14.32) to each sample sum separately. If there are error sources that are common to inspector and operator measurements—e.g., the use of a common measurement of the bulk amount of a transfer—then corrections for the common effects in the two totals are necessary in a manner completely analogous to those considered in Section 14.3.5 in connection with the combination of strata. For purposes of variance estimation, the variance of the difference between the sums x_{1i} and x_{2i} can be considered as just another case of the estimation of the variance of a linear combination of strata.

The variance of the estimate \hat{D}_i of the total discrepancy L_i present in a stratum obtained using Equation 17.19 includes a component due to the

variation of the sample total \hat{L}_i about its expected value $(n_i/N_i)L_i$ given by Equation 17.28 (Franklin, 1984). From Equations 17.19 and 17.30,

$$\begin{aligned} \text{Var}(\hat{D}_i) = \left[\frac{N_i}{n_i} \right]^2 & [\text{Var}(\hat{L}_i) + \text{Var}(x_{1i}) \\ & + \text{Var}(x_{2i}) - 2 \text{Cov}(x_{1i}, x_{2i})] \quad (17.32) \end{aligned}$$

Note that $\text{Var} \left(\sum_{k=1}^{n_i} \epsilon_{jik} \right)$ has been replaced by $\text{Var}(x_{ji})$ for $j = 1, 2$. This follows from Equations 17.22 and 17.23, regarding L_{iK} as a constant. From finite sampling theory, it follows that

$$\text{Var}(\hat{L}_i) = n_i \left[1 - \frac{n_i}{N_i} \right] \text{Var}(L_{iK}) \quad (17.33)$$

where

$$\text{Var}(L_{iK}) = \frac{\sum_{K=1}^{N_i} (L_{iK} - \bar{L}_i)^2}{N_i - 1} \quad (17.34)$$

This component of the variance of the differences arising from the sampling of discrepancies is an important consideration in testing and estimating the statistic \hat{D}_i , because only a test of the null hypothesis $L_i = 0$ is independent of the actual distribution of the discrepancies L_{iK} . When determining optimal tests, the distribution of the test statistic under the alternative hypothesis $L_i > 0$ will always depend on the assumed distribution of the individual discrepancies, and estimation procedures can only be optimized when the distribution is known. While the assumption of normality may be reasonable with respect to measurement error or other processing discrepancies present in individual items, it is not a very reasonable assumption in the case of discrepancies due to diversion. In most test procedures it is tacitly assumed either that the individual discrepancies are constant, in which case $\text{Var}(\hat{L}_i) = 0$, or that they are proportional to the content of the items and can, like the item contents, be assumed to be approximately normal.

Similar considerations apply to the variance of the statistic \hat{D} given by Equation 17.21. A direct application of Equation 2.123 yields

$$\text{Var}(\hat{D}) = \sum_{i=1}^{\ell} \text{Var}(\hat{D}_i) + 2 \sum_{i < j}^{\ell} c_i c_j \text{Cov}(\hat{D}_i, \hat{D}_j) \quad (17.35)$$

As is the case when the combination of stratum and component totals is considered in Section 14.3.5, covariances between estimates \hat{D}_i for different strata can occur both because of common error effects and because of the cancellation of identical results in two components of the inventory difference. It follows from Equation 17.19 that

$$\text{Cov}(\hat{D}_i, \hat{D}_j) = N_i N_j \text{Cov}(\bar{d}_i, \bar{d}_j)$$

From Equation 17.29, it follows that

$$n_i n_j \text{Cov}(\bar{d}_i, \bar{d}_j) = \text{Cov}(x_{1i}, x_{1j}) - \text{Cov}(x_{1i}, x_{2j}) - \text{Cov}(x_{2i}, x_{1j}) + \text{Cov}(x_{2i}, x_{2j})$$

so that

$$\begin{aligned} \text{Cov}(\hat{D}_i, \hat{D}_j) &= (N_i/n_i)(N_j/n_j) \\ &\quad \times [\text{Cov}(x_{1i}, x_{1j}) - \text{Cov}(x_{1i}, x_{2j}) - \text{Cov}(x_{2i}, x_{1j}) + \text{Cov}(x_{2i}, x_{2j})] \end{aligned}$$

Combining this result with Equation 17.32, Equation 17.35 becomes

$$\begin{aligned} \text{Var}(\hat{D}) &= \sum_{i=1}^{\ell} (N_i/n_i)^2 [\text{Var}(\hat{L}_i) + \text{Var}(x_{1i}) + \text{Var}(x_{2i}) - 2 \text{Cov}(x_{1i}, x_{2i})] \\ &\quad + 2 \sum_{i < j=1}^{\ell} c_i c_j (N_i/n_i)(N_j/n_j) [\text{Cov}(x_{1i}, x_{1j}) - \text{Cov}(x_{1i}, x_{2j}) \\ &\quad - \text{Cov}(x_{2i}, x_{1j}) + \text{Cov}(x_{2i}, x_{2j})] \end{aligned} \quad (17.36)$$

Because from Equations 17.21 and 17.19,

$$\hat{D} = \sum_{i=1}^{\ell} c_i N_i \bar{d}_i$$

it follows from Equation 17.29 that

$$\hat{D} = \sum_{i=1}^{\ell} c_i \left(\frac{N_i}{n_i} \right) (x_{1i} - x_{2i})$$

$$= \sum_{i=1}^{\ell} c_i \left(\frac{N_i}{n_i} \right) x_{1i} - \sum_{i=1}^{\ell} c_i \left(\frac{N_i}{n_i} \right) x_{2i} \quad (17.37)$$

The determination of that part of the variance of \hat{D} due to measurement error is a straightforward application of the methods of Section 14.3.5 to this linear combination of the 2ℓ sample sums x_{1i} and x_{2i} computed from the sampled items. Applying Equation 2.123 directly to Equation 17.37 accounts for all of the terms in Equation 17.36 except for the component $\sum_i (N_i/n_i)^2 \text{Var}(\hat{L}_i)$ due to the sampling of discrepancies. In this form, it can also be seen that, as in the case of determining the variances of inventory differences and their components, the measurement error contribution can be computed independently for each error source by writing the total error in each sample sum as a linear combination of independent error effects and substituting into Equation 17.37, so that \hat{D} is expressed as a linear combination of the independent error effects from each source.

17.3.2.3 The Variance of the Corrected Inventory Difference

Using the difference statistics and their estimated variances developed in the two preceding sections, the hypothesis that the total discrepancy L_i for an individual stratum, or the total discrepancy $L = \sum_{i=1}^{\ell} c_i L_i$ for a material balance, is zero can be tested against the alternative hypothesis of a nonzero discrepancy by the same procedures used for testing inventory differences and cumulative inventory differences in Section 16.2. The power of the testing procedure may vary widely depending on the true alternative.

For some purposes, it is more desirable to estimate a stratum total or an inventory difference which is corrected for possible discrepancies. In particular, the estimate $ID - \hat{D}$ is sometimes referred to as the "inspector's inventory difference." It follows from Equations 14.42 and 17.21 that

$$\begin{aligned} ID - \hat{D} &= \sum_{i=1}^{\ell} c_i x_i - \sum_{i=1}^{\ell} c_i D_i \\ &= \sum_{i=1}^{\ell} c_i (x_i - \hat{D}_i) \end{aligned} \quad (17.38)$$

where x_i is the sum of all N_i operator-reported values in stratum i . Thus, the inspector's estimate of the inventory difference is again a linear combination of the inspector's estimates of the stratum totals, and it will be sufficient to examine the variances (and possible covariances) of the stratum totals in order to obtain its variance.

The quantities x_i and \hat{D}_i in the corrected estimates of the stratum totals are not independent, since they both involve the n_i reported values of the operator for the items measured by the inspector. If $x'_i = x_i - x_{1i}$ designates the total reported amount in items not sampled, then from Equations 17.19 and 17.29,

$$x_i - \hat{D}_i = x'_i + \left(1 - \frac{N_i}{n_i}\right)x_{1i} + \frac{N_i}{n_i}x_{2i} \quad (17.39)$$

From this form, it is easy to see that as n_i approaches N_i , i.e., as the sample measured by the inspector approaches all the items in the stratum, $1 - N_i/n_i$ and x'_i both approach zero, N_i/n_i approaches one, and the corrected estimate approaches the total of the inspector measurements for all the items in the stratum.

From Equation 17.39, using the general formula for the variance of a linear combination of random variables given in Equation 2.123, $\text{Var}(x_i - \hat{D}_i)$ can be expressed as a linear combination of the variances and covariances of the totals x'_i , x_{1i} , and x_{2i} . As noted in the previous discussion of the variance of the difference statistic, the error effects associated with operator and inspector measurements usually come from completely independent sources, except when a common measurement is used by both. In such a case, the common error effects cancel identically from the differences between operator and inspector measurements, and the variability due to these common effects does not affect the variance of the total difference $n_i\bar{d}_i = x_{1i} - x_{2i}$. On the other hand, the operator error effects in the totals x'_i and x_{1i} are clearly from the same sources, and any of these which affect more than one measurement will presumably affect both sampled and unsampled items and lead to a positive covariance between the totals.

Because the estimates $\hat{V}\text{ar}(x_{1i})$ and $\hat{V}\text{ar}(x_{2i})$ and any necessary correction to their sum due to common effects will have been computed in calculating the estimated variance $\hat{V}\text{ar}(\hat{D}_i)$ of the difference statistic, it is more convenient to calculate the variance of the inspector's estimate from the expression

$$x_i - \hat{D}_i = x_i - \frac{N_i}{n_i}x_{1i} + \frac{N_i}{n_i}x_{2i} \quad (17.40)$$

Using Equation 2.123,

$$\text{Var}(x_i - \hat{D}_i) = \text{Var}(x_i) + \left(\frac{N_i}{n_i}\right)^2 \text{Var}(x_{1i})$$

$$\begin{aligned}
& + \left(\frac{N_i}{n_i} \right)^2 \text{Var}(x_{2i}) - 2 \left(\frac{N_i}{n_i} \right) \text{Cov}(x_i, x_{1i}) \\
& + 2 \left(\frac{N_i}{n_i} \right) \text{Cov}(x_i, x_{2i}) - 2 \left(\frac{N_i}{n_i} \right)^2 \text{Cov}(x_{1i}, x_{2i}) \quad (17.41)
\end{aligned}$$

Under the reasonable assumption that the errors in the inspector's measurements and the operator's declared measurements on non-sampled items are independent, $\text{Cov}(x_i, x_{2i}) = 0$. This, coupled with Equation 17.31, allows Equation 17.41 to be written

$$\text{Var}(x_i - \hat{D}_i) = \text{Var}(x_i) + N_i^2 \text{Var}(\bar{d}_i) - \frac{2N_i}{n_i} \text{Cov}(x_i, x_{1i}) \quad (17.42)$$

Because the sample total x_{1i} is the sum of a subset of the reported values included in the overall total x_i for the stratum, the measurement error sources contributing to both totals are identical. Hence, $\text{Var}(x_i)$ and $\text{Var}(x_{1i})$ differ only in whether the quantities μ and $\mu_{j(r)}$ in Equation 14.32 are based on the sums μ_i and $\mu_{ij(r)}$ for all items in the stratum or the sums μ_{1i} and $\mu_{1ij(r)}$ for items in the sample. It therefore follows from the direct method for calculating the covariance used in Equation 14.49 that

$$\text{Cov}(x_i, x_{1i}) = \mu_i \mu_{1i} \sum_r \theta_r^2 + \sum_r (\sum_{j(r)} \mu_{ij(r)} \mu_{1ij(r)}) \delta_r^2 / b_r \quad (17.43)$$

Assuming $\mu_{1ij(r)} = (n_i/N_i)\mu_{ij(r)}$ for each of the r sources of measurement error, it follows that $\mu_{1i} = (n_i/N_i)\mu_i$ and Equation 17.43 reduces to

$$\text{Cov}(x_i, x_{1i}) = (n_i/N_i)\text{Cov}(x_i, x_{1i}) \quad (17.44)$$

Because $\text{Var}(\hat{D}_i) = N_i^2 \text{Var}(\bar{d}_i)$, Equation 17.42 becomes

$$\text{Var}(x_i - \hat{D}_i) = \text{Var}(\hat{D}_i) - \text{Var}(x_i) \quad (17.45)$$

With the usual qualifications concerning any error effects common to the strata involved in the material balance, it follows from Equations 17.38, 17.21, and 14.42 that

$$\text{Var}(\text{ID} - \hat{D}) = \text{Var}(\hat{D}) - \text{Var}(\text{ID}) \quad (17.46)$$

While it may seem strange that the variance of a difference is the difference of the variances, remember that ID and \hat{D} are not independent and that the actual process is essentially one of replacing operator measurements by inspector measurements. $\text{Var}(ID - \hat{D})$ is a measure of the trust that the inspector can have in the reported inventory difference based on his measurements. The extreme cases are (1) no inspector measurements, in which case $\text{Var}(ID - \hat{D})$ becomes indeterminate, and (2) complete remeasurement of the entire material balance, in which case the inspector's measurement errors simply replace those of the operator. The primary difference in the latter case is simply that the inspector's measurements are free of any discrepancies which may have been deliberately introduced by the operator to conceal diversion. $\text{Var}(\hat{D})$, and hence $\text{Var}(ID - \hat{D})$, will be large relative to $\text{Var}(ID)$ if the sample is small or if the inspector's measurement variance is large compared to the operator's.

The assumption required to obtain Equation 17.44 from Equation 17.43 implies that the total amount of material in the sampled items must be the same constant fraction n_i/N_i of the total material in the stratum not only for the stratum as a whole, but for each of the a_r subsets of the N_i items in the stratum associated with every measurement source. In general, it is only true that with respect to the random sampling process for choosing the items to be verified, $E'(\mu_{ij(r)}) = (n_i/N_i)\mu_{ij(r)}$, so that Equations 17.45 and 17.46 only are true "on the average." Sufficient conditions that Equation 17.44 will always hold are (1) $\mu_{iK} = \mu_i/N_i$ for all K , and (2) the sources of measurement error are such that the error effects apply either to individual items or to the stratum total.

When the variances given by Equations 17.45 and 17.46 are used to form confidence limits on the inspector's estimate, $\text{Var}(\hat{D}_i)$ and $\text{Var}(\hat{D})$ must include the sampling variability given by Equation 17.33. Franklin (1984) has pointed out that if one defines discrepancies $L'_{iK} = L_{iK} + \epsilon_{iK}$ which include the operator's measurement error effects, then the variability in the corrected estimate is simply the combined variance associated with (1) sampling this finite population of discrepancies with expected value $L_i/N_i = \bar{L}_i$ and variance equal to the combined variance of the true discrepancies and the measurement errors, and (2) the inspector's measurement of the discrepancy.

Example 17.10 An inspection was carried out to verify the simplified material balance for the fuel fabrication facility considered in Example 14.6. The inspection activities are shown in Table 17.6.

TABLE 17.6
Data on Verification Activities

| | Stratum (i) | | | | | | |
|---|-------------|-----|----|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Number of batches sampled | 12 | 1 | 10 | 6 | 4 | 6 | 4 |
| Number of bulk measurements per sampled batch | 3 | 100 | 1 | 3 | 4 | 3 | 4 |
| Number of samples taken for chemical analysis per sampled batch | 2 | 24 | 1 | 2 | 3 | 2 | 3 |

The sampling was done in two stages: (1) a random selection of batches, and (2) a random selection of an equal number of items from each batch for verification of bulk amount and element concentration. Note that in Strata 2, 4, 5, 6, and 7, all of the arbitrarily defined batches were sampled. This proportional sampling process, along with the assumption of equal batch sizes, makes it possible to apply Equation 17.45 to each stratum.

It will be assumed that the verification activities for bulk measurement are carried out using the operator's scales but with independent calibration by the inspector. Thus, for measurement error Sources 1, 3, and 6, it will be assumed that both the common effect and the individual effects are independent for operator and inspector measurements, but that both sets of effects have the variances given in Example 14.6. The inspector's NDA measurements on a sample of waste batches will be assumed to be independent, but with error variance components θ_5' and δ_5' which are somewhat larger than those of the operator. The analyses for element concentration are all assumed to be carried out by the safeguards laboratory on independent samples. The values of the variance components are given in Table 17.7.

The total observed difference between sampled items for Stratum 1 is based on inspector and operator measurements on 36 containers of approximately 20 kg each, or a total of 720 kg. The contribution to the variance in the total difference from measurement error Source 1 is

$$\begin{aligned} \text{Var}_1(36\bar{d}_1) &= (720)^2(0.000439)^2 + 36(20)^2(0.000658)^2 \\ &= 0.106141(\text{kg})^2 \end{aligned}$$

TABLE 17.7
Relative Standard Deviations of
Inspector Measurements

| Source | Standard deviation | |
|--------|--------------------|-----------|
| | θ' | δ' |
| 1 | 0.000439 | 0.000658 |
| 2 | 0.000172 | 0.000433 |
| 3 | 0.000175 | 0.000877 |
| 4 | 0.000165 | 0.000822 |
| 5 | 0.0692 | 0.0923 |
| 6 | 0.00167 | 0.00250 |
| 7,8 | 0.00225 | 0.0285 |

for both operator and inspector. The operator contribution from the determination of element concentration is

$$\begin{aligned}\text{Var}_2(36\bar{d}_1) &= (720)^2(0.000571)^2 + 12(60)^2(0.000685)^2/5 \\ &= 0.173074(\text{kg})^2\end{aligned}$$

and the inspector contribution is

$$\begin{aligned}\text{Var}'_2(36\bar{d}_1) &= (720)^2(0.000172)^2 + 12(60)^2(0.000433)^2/2 \\ &= 0.019386(\text{kg})^2\end{aligned}$$

Hence, the variance of the total difference in the sampled items is

$$\text{Var}(36\bar{d}_1) = 0.404742(\text{kg})^2$$

and the variance of the difference statistic for Stratum 1 is

$$\begin{aligned}\text{Var}(\hat{D}_1) &= \left(\frac{N_1}{n_1} \right)^2 \text{Var}(n_1\bar{d}_1) \\ &= \left(\frac{12,000}{36} \right)^2 (0.404742) \\ &= 44,971.35(\text{kg})^2\end{aligned}$$

From Example 14.6, the variance in the reported total for Stratum 1 is

$$\begin{aligned}\text{Var}(x_1) &= \text{Var}_1(x) + \text{Var}_2(x) \\ &= 29,950.34(\text{kg})^2\end{aligned}$$

From Equation 17.45, the variance of the corrected estimate of the stratum total is

$$\text{Var}(x_1 - \hat{D}_1) = 15,021.01(\text{kg})^2$$

The fact that this variance is smaller than the operator estimate, even though based on a relatively small sample, is due to (1) the fact that for both inspector and operator the variance components associated with the error effects common to all measurements are dominant, and (2) in the corrected estimate, the error effects which are constant for all operator measurements are exactly replaced by the comparable inspector effects when the item amounts are equal or, more generally, when the total contribution of these error sources to the sample total is exactly n_i/N_i times the contribution to the stratum total. This can easily be seen from Equation 14.27 if it is assumed that each operator measurement contains an error $\bar{\mu}_i\beta_r'$ which is constant for all measurements in the stratum total x_i . This effect occurs $N_i - n_i$ times in the sum x_i' and n_i times in the sample total x_{1i} , so that the total contribution to the stratum total x_i is $[N_i - n_i + n_i(1 - N_i/n_i)]\bar{\mu}_i\beta_r' = 0$. Conversely, if $\bar{\mu}_i\beta_r'$ is the constant contribution due to the comparable inspector effect which occurs n_i times in the sample total x_{2i} , the contribution from this source is $n_i(N_i/n_i)\bar{\mu}_i\beta_r' = N_i\bar{\mu}_i\beta_r' = \mu\beta_r'$, which is the contribution that arises from applying the common error effect from Source r in the inspector's measurements to the total amount in the stratum. Table 17.8 shows the contributions from each type of effect to each variance.

TABLE 17.8
Variance Components

| Quantity estimated | Source of variance | |
|--------------------|--------------------|--------------------|
| | Common effects | Individual effects |
| x_1 | 29,880.69 | 69.65 |
| \hat{D}_1 | 42,685.46 | 2285.89 |
| $x_1 - \hat{D}_1$ | 12,804.77 | 2216.24 |

Stratum 2 has only one batch; it is assumed that the inspector, like the operator, applies a single pellet factor (based on the analysis of 24 samples) to each of the 100 pellet trays for which the bulk amount is verified. In this case, the contribution of both operator and inspector from Source 3 is

$$\begin{aligned}\text{Var}_3(100\bar{d}_2) &= (500)^2(0.000175)^2 + 100(5)^2(0.000877)^2 \\ &= 0.0095791 \text{ (kg)}^2\end{aligned}$$

and for the separate contributions from Source 4

$$\begin{aligned}\text{Var}_4(100\bar{d}_2) &= (500)^2[(0.000341)^2 + (0.000568)^2/240] \\ &= 0.029406 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}'_4(100\bar{d}_2) &= (500)^2[(0.000165)^2 + (0.000822)^2/24] \\ &= 0.013845 \text{ (kg)}^2\end{aligned}$$

so that

$$\text{Var}(100\bar{d}_2) = 0.062409 \text{ (kg)}^2$$

and

$$\begin{aligned}\text{Var}(\hat{D}_2) &= (47,760/100)^2 \text{Var}(100\bar{d}_2) \\ &= 14,235.65 \text{ (kg)}^2\end{aligned}$$

From Example 14.6, summing the contributions from Source 3 and Source 4,

$$\text{Var}(x_2) = 8454.95 \text{ (kg)}^2$$

so that

$$\text{Var}(x_2 - \hat{D}_2) = 5780.70 \text{ (kg)}^2$$

Stratum 3 involves only error Source 5, and so for the total difference in the 10 containers of solid waste randomly selected for verification by NDA measurement,

$$\begin{aligned}\text{Var}_5(10\bar{d}_3) &= (4.33)^2(0.0462)^2 + 10(0.433)^2(0.0577)^2 \\ &= 0.046260 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}'_5(10\bar{d}_3) &= (4.33)^2(0.0692)^2 + 10(0.433)^2(0.0923)^2 \\ &= 0.105754 (\text{kg})^2\end{aligned}$$

$$\text{Var}(10\bar{d}_3) = 0.152015 (\text{kg})^2$$

$$\begin{aligned}\text{Var}(\hat{D}_3) &= (2770/10)^2 \text{Var}(10\bar{d}_3) \\ &= 11,663.95 (\text{kg})^2\end{aligned}$$

From Example 14.6,

$$\text{Var}(x_3) = 3075.32 (\text{kg})^2$$

so that

$$\text{Var}(x_3 - \hat{D}_3) = 8588.63 (\text{kg})^2$$

Strata 4 and 6, the beginning and ending inventories of dirty scrap, have the same estimated variance in this example for operator and inspector. The contribution for both operator and inspector from error Source 6, again assuming the common element factor based on the analysis of two samples is applied to all containers in a batch, is

$$\begin{aligned}\text{Var}_6(18\bar{d}_4) &= (72)^2(0.00167)^2 + 18(4)^2(0.00250)^2 \\ &= 0.016258 (\text{kg})^2\end{aligned}$$

and the contributions from Source 7 are

$$\begin{aligned}\text{Var}_7(18\bar{d}_4) &= (72)^2[(0.00896)^2 + (0.03284)^2/10] \\ &= 0.975256 (\text{kg})^2\end{aligned}$$

$$\begin{aligned}\text{Var}'_7(18\bar{d}_4) &= (72)^2[(0.00225)^2 + (0.0285)^2/2] \\ &= 2.131596 (\text{kg})^2\end{aligned}$$

so that

$$\text{Var}(18\bar{d}_4) = 3.139368 (\text{kg})^2$$

and

$$\begin{aligned}\text{Var}(\hat{D}_4) &= (1800/18)^2 \text{Var}(18\bar{d}_4) \\ &= 31,393.68 (\text{kg})^2\end{aligned}$$

Because the common error effects from measurement Sources 6, 7, and 8 do not contribute to $\text{Var}(\hat{ID})$, these components were not computed for individual strata in Example 14.6. The contribution of these common error effects to the above estimate is 4713.39 (kg)^2 , so that the remaining component of $26,680.29 \text{ (kg)}^2$ is due to individual error effects. This corresponds to a total contribution of 931.91 (kg)^2 to $\text{Var}(x_4)$ from individual errors, and hence a contribution of $25,748.38 \text{ (kg)}^2$ from individual errors to $\text{Var}(x_4 - \hat{D}_4)$. This imbalance reflects the large variance in the individual analytical results for dirty scrap and the relatively small number of samples taken for verification of element content.

For Strata 5 and 7, the computations are

$$\begin{aligned}\text{Var}_6(16\bar{d}_5) &= (80)^2(0.00167)^2 + 16(5)^2(0.00250)^2 \\ &= 0.020349 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}_8(16\bar{d}_5) &= (80)^2[(0.01000)^2 + (0.04998)^2/12] \\ &= 1.972267 \text{ (kg)}^2\end{aligned}$$

$$\begin{aligned}\text{Var}'_8(16\bar{d}_5) &= (80)^2[(0.00225)^2 + (0.0285)^2/3] \\ &= 1.765200 \text{ (kg)}^2\end{aligned}$$

$$\text{Var}(16\bar{d}_5) = 3.778165 \text{ (kg)}^2$$

$$\begin{aligned}\text{Var}(\hat{D}_5) &= (800/16)^2 \text{Var}(16\bar{d}_5) \\ &= 9445.41 \text{ (kg)}^2\end{aligned}$$

The component of this variance due to individual effects is 7675.17 (kg)^2 , and because the corresponding contribution in Example 14.6 was 832.97 (kg)^2 , the contribution from the variance of the individual effect to $\text{Var}(x_4 - \hat{D}_4)$ is 6842.20 (kg)^2 .

TABLE 17.9

Summary of Variance Components (kg)²

| Stratum | ID | \hat{D} | ID- \hat{D} |
|---------------------|--------|-----------|---------------|
| 1 | 29,950 | 44,971 | 15,021 |
| 2 | 8,455 | 14,236 | 5,781 |
| 3 | 3,075 | 11,664 | 8,589 |
| 4† | 932 | 26,680 | 25,748 |
| 5† | 833 | 7,675 | 6,842 |
| 6† | 932 | 26,680 | 25,748 |
| 7† | 833 | 7,675 | 6,842 |
| Total | 45,010 | 139,581 | 94,571 |
| Standard error (kg) | 212 | 374 | 308 |

†Contributions from individual errors only.

The data for the entire material balance are summarized in Table 17.9, with the components that cancel from the variance of the estimate of ID eliminated. The greatest improvement in the variance of the corrected estimate would come from an improved estimate by the inspector of the element concentration in dirty scrap.

17.3.2.4 Variables Sampling Plans

In the last section, the inspection sample sizes were regarded as given and the resulting variance of the difference statistic was calculated. In this section, methods for determining the inspection sample sizes are developed.

A few words of background and perspective are appropriate before taking up the details. Virtually every text on sampling theory presents methods for determining variables sampling plans, and the plans have been widely used in some fields of application, particularly in sample surveys. However, these methods have not been widely practiced in safeguards applications. The reasons are primarily economic and logistical. In safeguards applications, variables samples are typically (although not necessarily) based on destructive chemical analysis: the inspector weighs items, takes samples for element or isotope determination, and then sends the samples to an offsite laboratory for analysis. These procedures are expensive (on the order of several hundred dollars per sample) and administratively complex, in part because of requirements for packaging and shipping radioactive materials. Another drawback is a lack of timeliness. Up to several months may elapse between the time of the inspection and the receipt of the analytical report from the offsite laboratory. For obvious reasons, it is more desirable to obtain and evaluate the results of the inspection quickly.

Because of these factors, the tendency in most inspection applications is to rely primarily on attributes testing or variables measurements using nondestructive techniques, taking only a small number of variables samples for calibrations, measurement control information, and other specialized purposes. In this section, therefore, only the basic principles of variables sampling plans will be presented, and no attempt will be made to give detailed guidance for applications. If the need arises for such guidance, the reader may consult such references as Jaech (1973) or the IAEA Safeguards Technical Manual (IAEA, 1988), which give full details.

As is usual in designing a sampling plan, it is first necessary to decide

1. The size of the difference that is of concern and is to be detected through the statistical sampling and testing.
2. The probability with which such a difference should be detected, if it exists (i.e., the power of the test).
3. The allowable probability of concluding that a difference exists when in fact it does not (i.e., the Type I error or false alarm probability).

The size of the difference that would be of concern is typically based on the concept of a significant quantity discussed earlier. In safeguards sampling plan calculations, this quantity is usually called the goal quantity and is denoted by the letter M . The power of the test, or one minus the Type II error, is conventionally chosen as some high value, say 0.90 or 0.95, although these values are not suitable in every case and should not be accepted uncritically. The Type I error probability is conventionally assigned a low value such as 0.05 and 0.10. Again, these conventional values should not be accepted uncritically.

The variance of the difference statistic is a function of the inspection sampling plan. A completely general treatment would be complex, because the variance depends on the sampling plan in a rather complicated manner, as is apparent from the discussion in Section 17.3.2.2. For purposes of exposition, a procedure for determining the optimum sample sizes for each stratum will be presented here. The procedure is based on the simplified assumptions of constant true item content and constant batch size associated with each replicated error source used to obtain Equation 14.34. The basic result is well known in sampling theory and is discussed in most sampling texts under the heading of optimum allocation (e.g., see Cochran, 1977; Hansen, Hurwitz, and Madow, 1953; or Kish, 1965), or sometimes as the Neyman allocation or Tschuprow-Neyman allocation after its originators (see Lehmann, 1983). In Section 8.3.4.2, a special case of the Neyman allocation is quoted in the context of stratified sampling to estimate a population mean or total. An analogy to the difference statistic can be drawn because the difference statistic estimates a (population) total difference between operator and inspector. However, there are details that make the analogy less than perfect.

When Equation 14.34 is used in connection with the procedures for computing the variance of \hat{D} discussed in Section 17.3.2.2, the components of variance due to each stratum are combined as described in Section 14.3.5. The linear combination, over all strata used to compute \hat{D} , of the components $\Sigma_r \theta_r^2$ corresponding to nonreplicated effects is independent of the sample size used to verify individual strata. For replicated effects, if the simplified approach is used, the constant size of the a_r subsets (batches) of items corresponding to the unique error effects from measurement source r is $N_{ir} = N_i/a_{ir}$, where N_i is the number of items and a_{ir} the number of independent measurement errors from source r in the i^{th} stratum. Making the additional assumption that the sampling is such that the number of sampled items containing each unique error effect is $n_{ir} = (n_i/N_i)N_{ir} = n_i/a_{ir}$, it follows that for each stratum i the variance component in Equation 14.34 correspondingly to the replicated effects can be written as $(1/n_i)\Sigma_r n_{ir} \delta_r^2/b_r$ for sampled items. Also, under the simplified assumptions, no replicated effects are common to operators and inspectors or to two strata, so that all the covariance terms of Equation 17.36 disap-

pear for those components that are dependent on n_i . Consequently, the variance of the difference statistic can generally be expressed as

$$\text{Var}(\hat{D}) = \tau^2 + \sum_{i=1}^k \frac{\tau_i^2}{n_i} \quad (17.47)$$

where

τ^2 = the sum of all the variance and covariance components associated with nonreplicated effects

$\frac{\tau_i^2}{n_i}$ = the variance component for stratum i dependent on n_i

$\tau_i^2 = \sum_r (n_{ir}/b_r)(\delta_{1r}^2 + \delta_{2r}^2)$

and δ_{1r}^2 and δ_{2r}^2 are the relative error variances of operator and inspector for (replicated) source r .

Because the difference statistic is assumed to be approximately normally distributed, the following probability statements can be written to express the design criteria:

$$\Pr\left\{\frac{\hat{D}}{\text{SE}(\hat{D})} > z_{1-\alpha} \mid E(\hat{D}) = 0\right\} = \alpha$$

$$\Pr\left\{\frac{\hat{D}}{\text{SE}(\hat{D})} > z_{1-\alpha} \mid E(\hat{D}) = M\right\} = 1 - \beta$$

where $M = \Sigma L_i > 0$ is some specified total loss. As discussed in Section 17.3.2.2, when M has a nonzero value, $\text{SE}(\hat{D})$ is a function of the nonzero component $\text{Var}(\hat{L}_i)$ included in the estimates of $\text{Var}(\hat{D}_i)$. The following simplified allocation plan is based only on the components τ_i^2 due to measurement error. The basic constraint on $\text{SE}(\hat{D})$ for fixed α , β , and M arising from these criteria is

$$(z_{1-\alpha} + z_{1-\beta})\text{SE}(\hat{D}) = M \quad (17.48)$$

Satisfying this constraint provides a target value for the overall variance $\text{Var}(\hat{D}) = [\text{SE}(\hat{D})]^2$. The next step is to determine how many samples should be taken in each stratum to achieve this target variance.

There is no unique solution unless an additional requirement is imposed. The essence of the Neyman allocation is to specify, as the additional requirement, that costs should be minimized. If the cost per item of a variables sample is the same for each stratum, then the total sample size for all strata should be minimized, subject to Equation 17.48. This leads to a sampling plan with the sample size in stratum i given by

$$n_i = \frac{\sum \tau_k}{V_T} \tau_i$$

where

$$\begin{aligned} V_T &= \text{the target value for the component of the variance of the difference statistic dependent on the } n_i \\ &= \text{Var}(\hat{D}) - \tau^2 \end{aligned}$$

In other words, the sample size is proportional to τ_i . The derivation is straightforward and utilizes Lagrange multipliers; an alternative proof based on the Cauchy-Schwarz inequality is sketched in Cochran (1977).

The rather simplified approach described here can become considerably more complicated in practical applications. However, the basic principles remain useful under a wide range of circumstances. Two cases illustrate the complexities that can arise. First, it may happen that no sampling plan exists that satisfies the specified design criteria. This will occur when the target variance needed to satisfy Equation 17.48 is smaller than τ^2 , the constant component of the variance. Even an infinite sample size cannot reduce the variance below this constant value, which is effectively a lower bound on the variance. The second possibility, which applies to sampling plan determinations generally, is that knowledge of the variance is required *a priori* when in fact the available information may be quite limited. In verification applications, in particular, the variance of the difference statistic may depend on the distribution of discrepancies, which is not known in advance. Franklin (1984) and the references therein discuss this and related problems.

REFERENCES

- ANSI, 1982, Generic Requirements for Auditing Nuclear Materials Safeguards Systems, ANSI N15.38-1982, American National Standards Institute, New York.
- Cochran, W. G., 1977, *Sampling Techniques*, 3rd Ed., John Wiley & Sons, Inc., New York.
- Committee on Basic Auditing Concepts, 1972, A Statement of Basic Auditing Concepts: A Review, *The Accounting Review*, Vol. 47, Suppl., The American Accounting Association, Sarasota, Florida.

- Deming, W. E., 1960, *Sample Design in Business Research*, John Wiley & Sons, Inc., New York.
- Franklin, M., 1984, The Sampling Distribution of the D Statistic, *J. Inst. Nucl. Mater. Manage. Proc.*, **25**: 135-142.
- Hansen, M. H., W. N. Hurwitz, and W. G. Madow, 1953, *Sample Survey Methods and Theory*, John Wiley & Sons, Inc., New York.
- IAEA, 1980, *IAEA Safeguards: Guidelines for States' Systems of Accounting for and Control of Nuclear Materials*, IAEA/SG/INF/2, International Atomic Energy Agency, Vienna.
- , 1988, *IAEA Safeguards Technical Manual, Part F*, International Atomic Energy Agency, Vienna.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, NTIS, Springfield, Virginia.
- Kendall, M., and A. Stuart, 1979, *The Advanced Theory of Statistics, Vol. 2*, 4th Ed., Macmillan Publishing Co., Inc., New York.
- Kish, L., 1965, *Survey Sampling*, John Wiley & Sons, Inc., New York.
- Lehmann, E. L., 1983, *Theory of Point Estimation*, John Wiley & Sons, Inc., New York.
- Sherr, T. S., 1972, *Attribute Sampling Inspection Procedure Based on the Hypergeometric Distribution*, WASH-1210, U.S. Atomic Energy Commission, Washington, D.C.
- Snedecor, G. W., and W. G. Cochran, 1980, *Statistical Methods*, 7th Ed., Iowa State University Press, Ames, Iowa.

Appendix A

Statistical Tables

APPENDIX A

STATISTICAL TABLES

| | Page |
|--|-------------|
| Table A1 Cumulative Binomial Distribution | 928 |
| Table A2 Cumulative Poisson Distribution | 934 |
| Table A3 Cumulative Standard Normal Distribution | 936 |
| Table A4 Quantile Values for the Chi-Square Distribution | 937 |
| Table A5 Quantile Values for Student's t Distribution | 939 |
| Table A6 Quantile Values for the F Distribution | 940 |
| Table A7 Pearson and Hartley Charts for the
Power of the F-Test | 943 |
| Table A8 Significant Ranges for Duncan's Multiple Range Test | 951 |
| Table A9 Critical Values for Standardized Residuals | 953 |
| Table A10 Durbin-Watson Critical Values | 955 |
| Table A11 Cumulative Probabilities of the
Total Number of Runs | 958 |
| Table A12 Quantiles of the Wilcoxon Signed Rank Test Statistic | 959 |
| Table A13 Quantiles of the Mann-Whitney Test Statistic | 961 |
| Table A14 Quantiles of the Squared Ranks Test Statistic | 966 |
| Table A15 Quantiles of the Spearman Test Statistic | 968 |
| Table A16 Quantiles of the Kendall Test Statistic | 969 |
| Table A17 Quantiles of the Kolmogorov Test Statistic | 971 |
| Table A18 Quantiles of the Lilliefors Test Statistic for Normality ... | 972 |
| Table A19 Coefficients a_i Used in the Shapiro-Wilk Test for
Normality | 973 |
| Table A20 Quantiles of the Shapiro-Wilk Test Statistic | 976 |
| Table A21 Critical Values for T_n When Standard Deviation
Is Calculated from the Same Sample | 977 |
| Table A22 Critical Values for w/s | 978 |
| Table A23 Critical Values for Simultaneously Testing the
Two Largest or Two Smallest Observations | 979 |
| Table A24 Critical Values for Testing $\sqrt{b_1}$ | 980 |
| Table A25 Critical Values for Testing b_2 | 980 |
| Table A26 Critical Values for T'_n When Standard Deviation
$s[\nu]$ Is Independent of Present Sample | 981 |
| Table A27 Critical Values of T'_1 and T'_n When the
Population Standard Deviation σ Is Known | 982 |
| Table A28 Critical Values for Screening Studentized Residuals | 983 |

| n | x | p = 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | |
|----|----|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 9 | 0 | 0.6302 | 0.3874 | 0.2316 | 0.1342 | 0.0751 | 0.0404 | 0.0207 | 0.0101 | 0.0046 | 0.0020 | 0.0008 | 0.0003 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.9288 | 0.7748 | 0.5995 | 0.4362 | 0.3003 | 0.1960 | 0.1211 | 0.0705 | 0.0385 | 0.0195 | 0.0091 | 0.0038 | 0.0014 | 0.0004 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.9916 | 0.9470 | 0.8591 | 0.7382 | 0.6007 | 0.4628 | 0.3373 | 0.2318 | 0.1495 | 0.0898 | 0.0498 | 0.0250 | 0.0112 | 0.0043 | 0.0013 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9994 | 0.9917 | 0.9661 | 0.9144 | 0.8343 | 0.7297 | 0.6089 | 0.4826 | 0.3614 | 0.2539 | 0.1648 | 0.0994 | 0.0536 | 0.0253 | 0.0100 | 0.0031 | 0.0006 | 0.0001 | 0.0000 | 0.0000 |
| | 4 | 1.0000 | 0.9991 | 0.9944 | 0.9804 | 0.9511 | 0.9012 | 0.8283 | 0.7334 | 0.6214 | 0.5000 | 0.3786 | 0.2666 | 0.1717 | 0.0988 | 0.0489 | 0.0196 | 0.0056 | 0.0009 | 0.0000 | 0.0000 |
| | 5 | 1.0000 | 0.9999 | 0.9994 | 0.9969 | 0.9900 | 0.9747 | 0.9464 | 0.9006 | 0.8342 | 0.7461 | 0.6386 | 0.5174 | 0.3911 | 0.2703 | 0.1657 | 0.0856 | 0.0339 | 0.0083 | 0.0006 | 0.0000 |
| | 6 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9987 | 0.9957 | 0.9888 | 0.9750 | 0.9502 | 0.9102 | 0.8505 | 0.7682 | 0.6627 | 0.5372 | 0.3993 | 0.2618 | 0.1409 | 0.0530 | 0.0084 | 0.0000 |
| | 7 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9996 | 0.9986 | 0.9962 | 0.9909 | 0.9805 | 0.9615 | 0.9295 | 0.8789 | 0.8040 | 0.6997 | 0.5638 | 0.4005 | 0.2252 | 0.0712 | 0.0000 |
| | 8 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9992 | 0.9980 | 0.9954 | 0.9899 | 0.9793 | 0.9596 | 0.9249 | 0.8658 | 0.7684 | 0.6126 | 0.3698 | 0.0000 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 10 | 0 | 0.5987 | 0.3487 | 0.1969 | 0.1074 | 0.0563 | 0.0282 | 0.0135 | 0.0060 | 0.0025 | 0.0010 | 0.0003 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.9139 | 0.7361 | 0.5443 | 0.3758 | 0.2440 | 0.1493 | 0.0860 | 0.0464 | 0.0233 | 0.0107 | 0.0045 | 0.0017 | 0.0005 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.9885 | 0.9298 | 0.8202 | 0.6778 | 0.5256 | 0.3828 | 0.2616 | 0.1673 | 0.0996 | 0.0547 | 0.0274 | 0.0123 | 0.0048 | 0.0016 | 0.0004 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9990 | 0.9872 | 0.9500 | 0.8791 | 0.7759 | 0.6496 | 0.5138 | 0.3823 | 0.2660 | 0.1719 | 0.1020 | 0.0548 | 0.0260 | 0.0106 | 0.0035 | 0.0009 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| | 4 | 0.9999 | 0.9984 | 0.9901 | 0.9672 | 0.9219 | 0.8497 | 0.7515 | 0.6331 | 0.5044 | 0.3770 | 0.2616 | 0.1662 | 0.0949 | 0.0473 | 0.0197 | 0.0064 | 0.0014 | 0.0001 | 0.0000 | 0.0000 |
| | 5 | 1.0000 | 0.9999 | 0.9986 | 0.9936 | 0.9803 | 0.9527 | 0.9051 | 0.8338 | 0.7384 | 0.6230 | 0.4956 | 0.3669 | 0.2485 | 0.1503 | 0.0781 | 0.0328 | 0.0099 | 0.0016 | 0.0001 | 0.0000 |
| | 6 | 1.0000 | 1.0000 | 0.9999 | 0.9991 | 0.9965 | 0.9894 | 0.9740 | 0.9452 | 0.8980 | 0.8281 | 0.7340 | 0.6177 | 0.4862 | 0.3504 | 0.2241 | 0.1209 | 0.0500 | 0.0128 | 0.0010 | 0.0000 |
| | 7 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9996 | 0.9984 | 0.9952 | 0.9877 | 0.9726 | 0.9453 | 0.9004 | 0.8327 | 0.7384 | 0.6172 | 0.4744 | 0.3222 | 0.1798 | 0.0702 | 0.0115 | 0.0000 |
| | 8 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9995 | 0.9983 | 0.9955 | 0.9893 | 0.9767 | 0.9536 | 0.9140 | 0.8507 | 0.7560 | 0.6242 | 0.4557 | 0.2639 | 0.0861 | 0.0000 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9990 | 0.9975 | 0.9940 | 0.9865 | 0.9718 | 0.9437 | 0.8926 | 0.8031 | 0.6513 | 0.4013 | 0.0000 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 11 | 0 | 0.5688 | 0.3138 | 0.1673 | 0.0859 | 0.0422 | 0.0198 | 0.0088 | 0.0036 | 0.0014 | 0.0005 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.8981 | 0.6974 | 0.4922 | 0.3221 | 0.1971 | 0.1130 | 0.0606 | 0.0302 | 0.0139 | 0.0059 | 0.0022 | 0.0007 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.9848 | 0.9104 | 0.7788 | 0.6174 | 0.4552 | 0.3127 | 0.2001 | 0.1189 | 0.0652 | 0.0327 | 0.0148 | 0.0059 | 0.0020 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9984 | 0.9815 | 0.9306 | 0.8389 | 0.7133 | 0.5696 | 0.4256 | 0.2963 | 0.1911 | 0.1133 | 0.0610 | 0.0293 | 0.0122 | 0.0043 | 0.0012 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 4 | 0.9999 | 0.9972 | 0.9841 | 0.9496 | 0.8854 | 0.7897 | 0.6683 | 0.5328 | 0.3917 | 0.2744 | 0.1738 | 0.0994 | 0.0501 | 0.0216 | 0.0076 | 0.0020 | 0.0003 | 0.0000 | 0.0000 | 0.0000 |
| | 5 | 1.0000 | 0.9997 | 0.9973 | 0.9883 | 0.9657 | 0.9218 | 0.8513 | 0.7535 | 0.6331 | 0.5000 | 0.3669 | 0.2465 | 0.1487 | 0.0782 | 0.0343 | 0.0117 | 0.0027 | 0.0003 | 0.0000 | 0.0000 |
| | 6 | 1.0000 | 1.0000 | 0.9997 | 0.9980 | 0.9924 | 0.9784 | 0.9499 | 0.9006 | 0.8262 | 0.7256 | 0.6029 | 0.4672 | 0.3317 | 0.2103 | 0.1146 | 0.0504 | 0.0159 | 0.0028 | 0.0001 | 0.0000 |
| | 7 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9988 | 0.9957 | 0.9878 | 0.9707 | 0.9390 | 0.8667 | 0.8089 | 0.7037 | 0.5744 | 0.4304 | 0.2867 | 0.1611 | 0.0694 | 0.0185 | 0.0016 | 0.0000 |
| | 8 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9980 | 0.9941 | 0.9852 | 0.9673 | 0.9348 | 0.8811 | 0.7999 | 0.6873 | 0.5448 | 0.3826 | 0.2212 | 0.0896 | 0.0152 | 0.0000 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9993 | 0.9978 | 0.9941 | 0.9861 | 0.9698 | 0.9394 | 0.8870 | 0.8029 | 0.6779 | 0.5078 | 0.3026 | 0.1019 | 0.0000 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9995 | 0.9986 | 0.9964 | 0.9912 | 0.9802 | 0.9578 | 0.9141 | 0.8327 | 0.6862 | 0.4312 | 0.0000 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 12 | 0 | 0.5404 | 0.2824 | 0.1422 | 0.0687 | 0.0317 | 0.0138 | 0.0057 | 0.0022 | 0.0008 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.8816 | 0.6590 | 0.4435 | 0.2749 | 0.1584 | 0.0850 | 0.0424 | 0.0196 | 0.0083 | 0.0032 | 0.0011 | 0.0003 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.9804 | 0.8891 | 0.7358 | 0.5583 | 0.3907 | 0.2528 | 0.1513 | 0.0834 | 0.0421 | 0.0193 | 0.0079 | 0.0028 | 0.0008 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9978 | 0.9744 | 0.9078 | 0.7946 | 0.6468 | 0.4925 | 0.3467 | 0.2253 | 0.1345 | 0.0730 | 0.0356 | 0.0153 | 0.0056 | 0.0017 | 0.0004 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 4 | 0.9998 | 0.9957 | 0.9761 | 0.9274 | 0.8424 | 0.7237 | 0.5833 | 0.4282 | 0.3044 | 0.1938 | 0.1117 | 0.0573 | 0.0255 | 0.0095 | 0.0028 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| | 5 | 1.0000 | 0.9995 | 0.9954 | 0.9806 | 0.9456 | 0.8822 | 0.7873 | 0.6652 | 0.5269 | 0.3872 | 0.2607 | 0.1582 | 0.0846 | 0.0386 | 0.0143 | 0.0039 | 0.0007 | 0.0001 | 0.0000 | 0.0000 |
| | 6 | 1.0000 | 0.9999 | 0.9993 | 0.9961 | 0.9857 | 0.9614 | 0.9154 | 0.8418 | 0.7393 | 0.6128 | 0.4731 | 0.3348 | 0.2127 | 0.1178 | 0.0544 | 0.0194 | 0.0046 | 0.0005 | 0.0000 | 0.0000 |
| | 7 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9972 | 0.9905 | 0.9745 | 0.9427 | 0.8883 | 0.8062 | 0.6956 | 0.5618 | 0.4167 | 0.2763 | 0.1576 | 0.0726 | 0.0239 | 0.0043 | 0.0002 | 0.0000 |
| | 8 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9996 | 0.9983 | 0.9944 | 0.9847 | 0.9644 | 0.9270 | 0.8655 | 0.7747 | 0.6533 | 0.5075 | 0.3512 | 0.2054 | 0.0922 | 0.0256 | 0.0022 | 0.0000 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9992 | 0.9972 | 0.9921 | 0.9807 | 0.9579 | 0.9166 | 0.8487 | 0.7472 | 0.6093 | 0.4417 | 0.2642 | 0.1109 | 0.0196 | 0.0000 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9989 | 0.9968 | 0.9917 | 0.9804 | 0.9576 | 0.9150 | 0.8416 | 0.7251 | 0.5655 | 0.3410 | 0.1184 | 0.0000 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9998 | 0.9992 | 0.9978 | 0.9943 | 0.9862 | 0.9683 | 0.9313 | 0.8578 | 0.7176 | 0.4596 | 0.0000 |
| | 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

(Table A1 continued on next page.)

| n | x | p = 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
|----|--------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 16 | 0 | 0.4401 | 0.1853 | 0.0743 | 0.0281 | 0.0100 | 0.0033 | 0.0010 | 0.0003 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.8108 | 0.5147 | 0.2839 | 0.1407 | 0.0635 | 0.0261 | 0.0098 | 0.0033 | 0.0010 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 0.9571 | 0.7892 | 0.5614 | 0.3518 | 0.1971 | 0.0994 | 0.0451 | 0.0183 | 0.0066 | 0.0021 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3 | 0.9930 | 0.9316 | 0.7899 | 0.5981 | 0.4050 | 0.2459 | 0.1339 | 0.0651 | 0.0281 | 0.0106 | 0.0035 | 0.0009 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | 0.9991 | 0.9830 | 0.9209 | 0.7982 | 0.6302 | 0.4499 | 0.2892 | 0.1666 | 0.0853 | 0.0384 | 0.0149 | 0.0049 | 0.0013 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5 | 0.9999 | 0.9967 | 0.9765 | 0.9183 | 0.8103 | 0.6598 | 0.4900 | 0.3288 | 0.1976 | 0.1051 | 0.0486 | 0.0191 | 0.0062 | 0.0016 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6 | 1.0000 | 0.9995 | 0.9944 | 0.9733 | 0.9204 | 0.8247 | 0.6881 | 0.5272 | 0.3660 | 0.2272 | 0.1241 | 0.0583 | 0.0229 | 0.0071 | 0.0016 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 7 | 1.0000 | 0.9999 | 0.9989 | 0.9930 | 0.9729 | 0.9256 | 0.8406 | 0.7161 | 0.5629 | 0.4018 | 0.2559 | 0.1423 | 0.0671 | 0.0257 | 0.0075 | 0.0015 | 0.0002 | 0.0000 | 0.0000 | 0.0000 |
| 8 | 1.0000 | 1.0000 | 0.9998 | 0.9985 | 0.9925 | 0.9743 | 0.9329 | 0.8577 | 0.7441 | 0.5982 | 0.4371 | 0.2839 | 0.1594 | 0.0744 | 0.0271 | 0.0070 | 0.0011 | 0.0001 | 0.0000 | 0.0000 |
| 9 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9984 | 0.9929 | 0.9771 | 0.9417 | 0.8759 | 0.7728 | 0.6340 | 0.4728 | 0.3119 | 0.1753 | 0.0796 | 0.0267 | 0.0056 | 0.0005 | 0.0001 | 0.0000 |
| 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9984 | 0.9938 | 0.9809 | 0.9514 | 0.8949 | 0.8024 | 0.6712 | 0.5100 | 0.3402 | 0.1897 | 0.0817 | 0.0235 | 0.0033 | 0.0001 | 0.0000 |
| 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9987 | 0.9951 | 0.9851 | 0.9616 | 0.9147 | 0.8334 | 0.7108 | 0.5501 | 0.3698 | 0.2018 | 0.0791 | 0.0170 | 0.0009 | 0.0000 |
| 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9991 | 0.9965 | 0.9894 | 0.9719 | 0.9349 | 0.8661 | 0.7541 | 0.5950 | 0.4019 | 0.2101 | 0.0684 | 0.0070 | 0.0000 |
| 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9979 | 0.9934 | 0.9817 | 0.9549 | 0.9006 | 0.8729 | 0.8482 | 0.4386 | 0.2108 | 0.0429 | 0.0000 |
| 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9990 | 0.9967 | 0.9902 | 0.9739 | 0.9365 | 0.8593 | 0.7161 | 0.4853 | 0.1892 | 0.0000 |
| 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9990 | 0.9967 | 0.9900 | 0.9719 | 0.9257 | 0.8147 | 0.5599 | 0.0000 |
| 16 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 17 | 0 | 0.4181 | 0.1668 | 0.0631 | 0.0225 | 0.0075 | 0.0023 | 0.0007 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.7922 | 0.4818 | 0.2525 | 0.1182 | 0.0501 | 0.0193 | 0.0067 | 0.0021 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 0.9497 | 0.7618 | 0.5198 | 0.3096 | 0.1637 | 0.0774 | 0.0327 | 0.0123 | 0.0041 | 0.0012 | 0.0003 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3 | 0.9912 | 0.9174 | 0.7556 | 0.5489 | 0.3530 | 0.2019 | 0.1028 | 0.0464 | 0.0184 | 0.0064 | 0.0019 | 0.0005 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | 0.9988 | 0.9779 | 0.9013 | 0.7582 | 0.5739 | 0.3887 | 0.2348 | 0.1260 | 0.0596 | 0.0245 | 0.0086 | 0.0025 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5 | 0.9999 | 0.9953 | 0.9681 | 0.8943 | 0.7653 | 0.5968 | 0.4197 | 0.2639 | 0.1471 | 0.0717 | 0.0301 | 0.0106 | 0.0030 | 0.0007 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6 | 1.0000 | 0.9992 | 0.9917 | 0.9623 | 0.8929 | 0.7752 | 0.6188 | 0.4478 | 0.2902 | 0.1662 | 0.0826 | 0.0348 | 0.0120 | 0.0032 | 0.0006 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 7 | 1.0000 | 0.9999 | 0.9983 | 0.9891 | 0.9598 | 0.8954 | 0.7872 | 0.6405 | 0.4743 | 0.3145 | 0.1834 | 0.0919 | 0.0383 | 0.0127 | 0.0031 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 8 | 1.0000 | 1.0000 | 0.9997 | 0.9974 | 0.9876 | 0.9597 | 0.9006 | 0.8011 | 0.6626 | 0.5000 | 0.3374 | 0.1989 | 0.0994 | 0.0403 | 0.0124 | 0.0026 | 0.0003 | 0.0000 | 0.0000 | 0.0000 |
| 9 | 1.0000 | 1.0000 | 1.0000 | 0.9995 | 0.9969 | 0.9873 | 0.9617 | 0.9081 | 0.8166 | 0.6955 | 0.5257 | 0.3595 | 0.2128 | 0.1046 | 0.0402 | 0.0109 | 0.0017 | 0.0001 | 0.0000 | 0.0000 |
| 10 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9968 | 0.9880 | 0.9652 | 0.9174 | 0.8338 | 0.7098 | 0.5522 | 0.3812 | 0.2248 | 0.1071 | 0.0377 | 0.0083 | 0.0008 | 0.0000 | 0.0000 |
| 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9993 | 0.9970 | 0.9894 | 0.9699 | 0.9283 | 0.8529 | 0.7361 | 0.5803 | 0.4032 | 0.2347 | 0.1057 | 0.0319 | 0.0047 | 0.0001 | 0.0000 |
| 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9975 | 0.9914 | 0.9755 | 0.9404 | 0.8740 | 0.7652 | 0.6113 | 0.4261 | 0.2418 | 0.0987 | 0.0221 | 0.0012 | 0.0000 |
| 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9995 | 0.9981 | 0.9936 | 0.9816 | 0.9536 | 0.8972 | 0.7981 | 0.6470 | 0.4511 | 0.2444 | 0.0826 | 0.0088 | 0.0000 |
| 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9988 | 0.9959 | 0.9877 | 0.9673 | 0.9226 | 0.8363 | 0.6904 | 0.4802 | 0.2382 | 0.0503 | 0.0000 |
| 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9994 | 0.9979 | 0.9933 | 0.9807 | 0.9499 | 0.8818 | 0.7475 | 0.5182 | 0.2078 |
| 16 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9993 | 0.9977 | 0.9925 | 0.9775 | 0.9369 | 0.8332 | 0.5819 |
| 17 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

(Table A1 continued on next page.)

| n | x | p = 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
|-----|--------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 20 | 0 | 0.3585 | 0.1216 | 0.0388 | 0.0115 | 0.0032 | 0.0008 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.7358 | 0.3917 | 0.1756 | 0.0692 | 0.0243 | 0.0076 | 0.0021 | 0.0005 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.9245 | 0.6769 | 0.4049 | 0.2061 | 0.0913 | 0.0355 | 0.0121 | 0.0036 | 0.0009 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9841 | 0.8670 | 0.6477 | 0.4114 | 0.2252 | 0.1071 | 0.0444 | 0.0160 | 0.0049 | 0.0013 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 4 | 0.9974 | 0.9568 | 0.8298 | 0.6296 | 0.4148 | 0.2375 | 0.1182 | 0.0510 | 0.0189 | 0.0059 | 0.0015 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 5 | 0.9997 | 0.9887 | 0.9327 | 0.8042 | 0.6172 | 0.4164 | 0.2454 | 0.1256 | 0.0553 | 0.0207 | 0.0064 | 0.0016 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 6 | 1.0000 | 0.9976 | 0.9781 | 0.9133 | 0.7858 | 0.6080 | 0.4166 | 0.2500 | 0.1299 | 0.0577 | 0.0214 | 0.0065 | 0.0015 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 7 | 1.0000 | 0.9996 | 0.9941 | 0.9679 | 0.8982 | 0.7723 | 0.6010 | 0.4159 | 0.2520 | 0.1316 | 0.0580 | 0.0210 | 0.0060 | 0.0013 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 8 | 1.0000 | 0.9999 | 0.9987 | 0.9900 | 0.9591 | 0.8867 | 0.7624 | 0.5956 | 0.4143 | 0.2517 | 0.1308 | 0.0565 | 0.0196 | 0.0051 | 0.0009 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| | 9 | 1.0000 | 1.0000 | 0.9998 | 0.9974 | 0.9861 | 0.9520 | 0.8782 | 0.7553 | 0.5914 | 0.4119 | 0.2493 | 0.1275 | 0.0532 | 0.0171 | 0.0039 | 0.0006 | 0.0000 | 0.0000 | 0.0000 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 0.9994 | 0.9961 | 0.9829 | 0.9468 | 0.8725 | 0.7507 | 0.5881 | 0.4086 | 0.2447 | 0.1218 | 0.0480 | 0.0139 | 0.0026 | 0.0002 | 0.0000 | 0.0000 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9991 | 0.9949 | 0.9804 | 0.9435 | 0.8692 | 0.7483 | 0.5857 | 0.4044 | 0.2376 | 0.1133 | 0.0409 | 0.0100 | 0.0013 | 0.0001 | 0.0000 |
| | 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9987 | 0.9940 | 0.9790 | 0.9420 | 0.8684 | 0.7480 | 0.5841 | 0.3990 | 0.2277 | 0.1018 | 0.0321 | 0.0059 | 0.0004 | 0.0000 |
| | 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9985 | 0.9935 | 0.9786 | 0.9423 | 0.8701 | 0.7500 | 0.5834 | 0.3920 | 0.2142 | 0.0867 | 0.0219 | 0.0024 | 0.0000 |
| | 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9984 | 0.9936 | 0.9793 | 0.9447 | 0.8744 | 0.7546 | 0.5836 | 0.3828 | 0.1958 | 0.0673 | 0.0113 | 0.0003 |
| | 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9985 | 0.9941 | 0.9811 | 0.9490 | 0.8818 | 0.7625 | 0.5852 | 0.3704 | 0.1702 | 0.0432 | 0.0026 |
| | 16 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9987 | 0.9951 | 0.9840 | 0.9556 | 0.8929 | 0.7748 | 0.5886 | 0.3523 | 0.1330 | 0.0159 |
| | 17 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9991 | 0.9964 | 0.9879 | 0.9645 | 0.9087 | 0.7939 | 0.5951 | 0.3231 | 0.0755 |
| | 18 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9995 | 0.9979 | 0.9924 | 0.9757 | 0.9308 | 0.8244 | 0.6083 | 0.2642 |
| | 19 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9992 | 0.9968 | 0.9885 | 0.9612 | 0.8784 | 0.6415 |
| 20† | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

*Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © 1980, John Wiley & Sons, Inc., Reprinted by permission of John Wiley & Sons, Inc., and the author.

†X has the binomial distribution with parameters n and p. The entries are the values of $p(X \leq x) = \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i}$, for p ranging from 0.5 to 0.95.

‡For n larger than 20, refer to Sections 2.4.3 and 2.5.2.2, which discuss the criteria for applying the Poisson and Normal approximations to the binomial.

| μ^x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2.6 | 0.074 | 0.267 | 0.518 | 0.736 | 0.877 | 0.951 | 0.983 | 0.995 | 0.999 | | | | | | | | | | |
| 2.7 | 0.067 | 0.249 | 0.494 | 0.714 | 0.863 | 0.943 | 0.979 | 0.993 | 0.998 | 0.999 | | | | | | | | | |
| 2.8 | 0.061 | 0.231 | 0.469 | 0.692 | 0.848 | 0.935 | 0.976 | 0.992 | 0.998 | 0.999 | | | | | | | | | |
| 2.9 | 0.055 | 0.215 | 0.446 | 0.670 | 0.832 | 0.926 | 0.971 | 0.990 | 0.997 | 0.999 | | | | | | | | | |
| 3.0 | 0.050 | 0.199 | 0.423 | 0.647 | 0.815 | 0.916 | 0.966 | 0.988 | 0.996 | 0.999 | | | | | | | | | |
| 3.2 | 0.041 | 0.171 | 0.380 | 0.603 | 0.781 | 0.896 | 0.955 | 0.983 | 0.994 | 0.998 | | | | | | | | | |
| 3.4 | 0.033 | 0.147 | 0.340 | 0.558 | 0.744 | 0.871 | 0.942 | 0.977 | 0.992 | 0.997 | 0.999 | | | | | | | | |
| 3.6 | 0.027 | 0.126 | 0.303 | 0.515 | 0.706 | 0.844 | 0.927 | 0.969 | 0.988 | 0.996 | 0.999 | | | | | | | | |
| 3.8 | 0.022 | 0.107 | 0.269 | 0.473 | 0.668 | 0.816 | 0.909 | 0.960 | 0.984 | 0.994 | 0.998 | 0.999 | | | | | | | |
| 4.0 | 0.018 | 0.092 | 0.238 | 0.433 | 0.629 | 0.785 | 0.889 | 0.949 | 0.979 | 0.992 | 0.997 | 0.999 | | | | | | | |
| 4.2 | 0.015 | 0.078 | 0.210 | 0.395 | 0.590 | 0.753 | 0.867 | 0.936 | 0.972 | 0.989 | 0.996 | 0.999 | | | | | | | |
| 4.4 | 0.012 | 0.066 | 0.185 | 0.359 | 0.551 | 0.720 | 0.844 | 0.921 | 0.964 | 0.985 | 0.994 | 0.998 | 0.999 | | | | | | |
| 4.6 | 0.010 | 0.056 | 0.163 | 0.326 | 0.513 | 0.686 | 0.818 | 0.905 | 0.955 | 0.980 | 0.992 | 0.997 | 0.999 | | | | | | |
| 4.8 | 0.008 | 0.048 | 0.143 | 0.294 | 0.476 | 0.651 | 0.791 | 0.887 | 0.944 | 0.975 | 0.990 | 0.996 | 0.999 | | | | | | |
| 5.0 | 0.007 | 0.040 | 0.125 | 0.265 | 0.440 | 0.616 | 0.762 | 0.867 | 0.932 | 0.968 | 0.986 | 0.995 | 0.998 | 0.999 | | | | | |
| 5.2 | 0.006 | 0.034 | 0.109 | 0.238 | 0.406 | 0.581 | 0.732 | 0.845 | 0.918 | 0.960 | 0.982 | 0.993 | 0.997 | 0.999 | | | | | |
| 5.4 | 0.005 | 0.029 | 0.095 | 0.213 | 0.373 | 0.546 | 0.702 | 0.822 | 0.903 | 0.951 | 0.977 | 0.990 | 0.996 | 0.999 | | | | | |
| 5.6 | 0.004 | 0.024 | 0.082 | 0.191 | 0.342 | 0.512 | 0.670 | 0.797 | 0.886 | 0.941 | 0.972 | 0.988 | 0.995 | 0.998 | 0.999 | | | | |
| 5.8 | 0.003 | 0.021 | 0.072 | 0.170 | 0.313 | 0.478 | 0.638 | 0.771 | 0.867 | 0.929 | 0.965 | 0.984 | 0.993 | 0.997 | 0.999 | | | | |
| 6.0 | 0.002 | 0.017 | 0.062 | 0.151 | 0.285 | 0.446 | 0.606 | 0.744 | 0.847 | 0.916 | 0.957 | 0.980 | 0.991 | 0.996 | 0.999 | 0.999 | | | |
| 6.2 | 0.002 | 0.015 | 0.054 | 0.134 | 0.259 | 0.414 | 0.574 | 0.716 | 0.826 | 0.902 | 0.949 | 0.975 | 0.989 | 0.995 | 0.998 | 0.999 | | | |
| 6.4 | 0.002 | 0.012 | 0.046 | 0.119 | 0.235 | 0.384 | 0.542 | 0.687 | 0.803 | 0.886 | 0.939 | 0.969 | 0.986 | 0.994 | 0.997 | 0.999 | | | |
| 6.6 | 0.001 | 0.010 | 0.040 | 0.105 | 0.213 | 0.355 | 0.511 | 0.658 | 0.780 | 0.869 | 0.927 | 0.963 | 0.982 | 0.992 | 0.997 | 0.999 | 0.999 | | |
| 6.8 | 0.001 | 0.009 | 0.034 | 0.093 | 0.192 | 0.327 | 0.480 | 0.628 | 0.755 | 0.850 | 0.915 | 0.955 | 0.978 | 0.990 | 0.996 | 0.998 | 0.999 | | |
| 7.0 | 0.001 | 0.007 | 0.030 | 0.082 | 0.173 | 0.301 | 0.450 | 0.599 | 0.729 | 0.830 | 0.901 | 0.947 | 0.973 | 0.987 | 0.994 | 0.998 | 0.999 | | |
| 7.2 | 0.001 | 0.006 | 0.025 | 0.072 | 0.156 | 0.276 | 0.420 | 0.569 | 0.703 | 0.810 | 0.887 | 0.937 | 0.967 | 0.984 | 0.993 | 0.997 | 0.999 | 0.999 | |
| 7.4 | 0.001 | 0.005 | 0.022 | 0.063 | 0.140 | 0.253 | 0.392 | 0.539 | 0.676 | 0.788 | 0.871 | 0.926 | 0.961 | 0.980 | 0.991 | 0.996 | 0.998 | 0.999 | |
| 7.6 | 0.001 | 0.004 | 0.019 | 0.055 | 0.125 | 0.231 | 0.365 | 0.510 | 0.648 | 0.765 | 0.854 | 0.915 | 0.954 | 0.976 | 0.989 | 0.995 | 0.998 | 0.999 | |
| 7.8 | | 0.004 | 0.016 | 0.048 | 0.112 | 0.210 | 0.338 | 0.481 | 0.620 | 0.741 | 0.835 | 0.902 | 0.945 | 0.971 | 0.986 | 0.993 | 0.997 | 0.999 | |
| 8.0 | | 0.003 | 0.014 | 0.042 | 0.100 | 0.191 | 0.313 | 0.453 | 0.593 | 0.717 | 0.816 | 0.888 | 0.936 | 0.966 | 0.983 | 0.992 | 0.996 | 0.998 | 0.999 |
| 8.2 | | 0.003 | 0.012 | 0.037 | 0.089 | 0.174 | 0.290 | 0.425 | 0.565 | 0.692 | 0.796 | 0.873 | 0.926 | 0.960 | 0.979 | 0.990 | 0.995 | 0.998 | 0.999 |
| 8.4 | | 0.002 | 0.010 | 0.032 | 0.079 | 0.157 | 0.267 | 0.399 | 0.537 | 0.666 | 0.774 | 0.857 | 0.915 | 0.952 | 0.975 | 0.987 | 0.994 | 0.997 | 0.999 |

*Ostle, Bernard, 1975, *Statistics in Research*, 3rd Ed., © 1975 by Iowa State University Press, Ames, Iowa, and E. C. Molina, 1947, *Poisson's Exponential Binomial Limit*, © 1947 (renewed 1970), by D. Van Nostrand, Inc. Reprinted with permission of the Iowa State University Press, B. Ostle, and Wadsworth Publishing Co., Inc., Belmont, California (current copyright holder for the Molina publication).

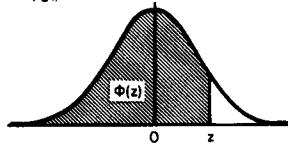
†Entries in the table are values of $F(x)$ where

$$F(x) = P(c \leq x) = \sum_{c=0}^x e^{-\mu} \mu^c / c!$$

Blank spaces to the right of the last entry in any row of the table may be read as 1; blank spaces to the left of the first entry in any row of the table may be read as 0.

TABLE A3
Cumulative Standard Normal Distribution*

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

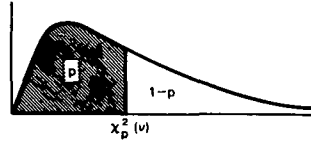


| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.54379 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57534 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62551 | 0.62930 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68438 | 0.68793 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.70540 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.72240 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.75490 |
| 0.7 | 0.75803 | 0.76115 | 0.76424 | 0.76730 | 0.77035 | 0.77337 | 0.77637 | 0.77935 | 0.78230 | 0.78523 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79954 | 0.80234 | 0.80510 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 | 0.82894 | 0.83147 | 0.83397 | 0.83646 | 0.83891 |
| 1.0 | 0.84134 | 0.84375 | 0.84613 | 0.84849 | 0.85083 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.86433 | 0.86650 | 0.86864 | 0.87076 | 0.87285 | 0.87493 | 0.87697 | 0.87900 | 0.88100 | 0.88297 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 | 0.89435 | 0.89616 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.90320 | 0.90490 | 0.90658 | 0.90824 | 0.90988 | 0.91149 | 0.91308 | 0.91465 | 0.91621 | 0.91773 |
| 1.4 | 0.91924 | 0.92073 | 0.92219 | 0.92364 | 0.92506 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |
| 1.5 | 0.93319 | 0.93448 | 0.93574 | 0.93699 | 0.93822 | 0.93943 | 0.94062 | 0.94179 | 0.94295 | 0.94408 |
| 1.6 | 0.94520 | 0.94630 | 0.94738 | 0.94845 | 0.94950 | 0.95053 | 0.95154 | 0.95254 | 0.95352 | 0.95448 |
| 1.7 | 0.95543 | 0.95637 | 0.95728 | 0.95818 | 0.95907 | 0.95994 | 0.96080 | 0.96164 | 0.96246 | 0.96327 |
| 1.8 | 0.96407 | 0.96485 | 0.96562 | 0.96637 | 0.96711 | 0.96784 | 0.96856 | 0.96926 | 0.96995 | 0.97062 |
| 1.9 | 0.97128 | 0.97193 | 0.97257 | 0.97320 | 0.97381 | 0.97441 | 0.97500 | 0.97558 | 0.97615 | 0.97670 |
| 2.0 | 0.97725 | 0.97778 | 0.97831 | 0.97882 | 0.97932 | 0.97982 | 0.98030 | 0.98077 | 0.98124 | 0.98169 |
| 2.1 | 0.98214 | 0.98257 | 0.98300 | 0.98341 | 0.98382 | 0.98422 | 0.98461 | 0.98500 | 0.98537 | 0.98574 |
| 2.2 | 0.98610 | 0.98645 | 0.98679 | 0.98713 | 0.98745 | 0.98778 | 0.98809 | 0.98840 | 0.98870 | 0.98899 |
| 2.3 | 0.98928 | 0.98956 | 0.98983 | 0.99010 | 0.99036 | 0.99061 | 0.99086 | 0.99111 | 0.99134 | 0.99158 |
| 2.4 | 0.99180 | 0.99202 | 0.99224 | 0.99245 | 0.99266 | 0.99286 | 0.99305 | 0.99324 | 0.99343 | 0.99361 |
| 2.5 | 0.99379 | 0.99396 | 0.99413 | 0.99430 | 0.99446 | 0.99461 | 0.99477 | 0.99492 | 0.99506 | 0.99520 |
| 2.6 | 0.99534 | 0.99547 | 0.99560 | 0.99573 | 0.99585 | 0.99598 | 0.99609 | 0.99621 | 0.99632 | 0.99643 |
| 2.7 | 0.99653 | 0.99664 | 0.99674 | 0.99683 | 0.99693 | 0.99702 | 0.99711 | 0.99720 | 0.99728 | 0.99736 |
| 2.8 | 0.99744 | 0.99752 | 0.99760 | 0.99767 | 0.99774 | 0.99781 | 0.99788 | 0.99795 | 0.99801 | 0.99807 |
| 2.9 | 0.99813 | 0.99819 | 0.99825 | 0.99831 | 0.99836 | 0.99841 | 0.99846 | 0.99851 | 0.99856 | 0.99861 |
| 3.0 | 0.99865 | 0.99869 | 0.99874 | 0.99878 | 0.99882 | 0.99886 | 0.99889 | 0.99893 | 0.99897 | 0.99900 |
| 3.1 | 0.99903 | 0.99906 | 0.99910 | 0.99913 | 0.99916 | 0.99918 | 0.99921 | 0.99924 | 0.99926 | 0.99929 |
| 3.2 | 0.99931 | 0.99934 | 0.99936 | 0.99938 | 0.99940 | 0.99942 | 0.99944 | 0.99946 | 0.99948 | 0.99950 |
| 3.3 | 0.99952 | 0.99953 | 0.99955 | 0.99957 | 0.99958 | 0.99960 | 0.99961 | 0.99962 | 0.99964 | 0.99965 |
| 3.4 | 0.99966 | 0.99968 | 0.99969 | 0.99970 | 0.99971 | 0.99972 | 0.99973 | 0.99974 | 0.99975 | 0.99976 |
| 3.5 | 0.99977 | 0.99978 | 0.99978 | 0.99979 | 0.99980 | 0.99981 | 0.99981 | 0.99982 | 0.99983 | 0.99983 |
| 3.6 | 0.99984 | 0.99985 | 0.99985 | 0.99986 | 0.99986 | 0.99987 | 0.99987 | 0.99988 | 0.99988 | 0.99989 |
| 3.7 | 0.99989 | 0.99990 | 0.99990 | 0.99990 | 0.99991 | 0.99991 | 0.99992 | 0.99992 | 0.99992 | 0.99992 |
| 3.8 | 0.99993 | 0.99993 | 0.99993 | 0.99994 | 0.99994 | 0.99994 | 0.99994 | 0.99995 | 0.99995 | 0.99995 |
| 3.9 | 0.99995 | 0.99995 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99997 | 0.99997 |

*Hines, W. W., and D. C. Montgomery, 1980, *Probability and Statistics in Engineering and Management Science*, 2nd Ed. (Appendix Table II), © 1980, John Wiley & Sons, Inc. Reprinted by permission of the publisher.

TABLE A4

Quantile Values $\chi_p^2(\nu)$ for the Chi-Square Distribution with ν Degrees of Freedom*



| ν | p | | | | | | | | | | | | | | | | |
|-------|------------|----------|----------|----------|--------|--------|-------|-------|-------|-------|------|------|------|------|-------|------|-------|
| | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
| 1 | 0.00000393 | 0.000157 | 0.000982 | 0.000393 | 0.0158 | 0.0642 | 0.146 | 0.275 | 0.455 | 0.708 | 1.07 | 1.64 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 0.446 | 0.713 | 1.02 | 1.39 | 1.83 | 2.41 | 3.22 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.0717 | 0.115 | 0.216 | 0.352 | 0.584 | 1.00 | 1.42 | 1.87 | 2.37 | 2.95 | 3.67 | 4.64 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.06 | 1.65 | 2.19 | 2.75 | 3.36 | 4.04 | 4.88 | 5.99 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.15 | 1.61 | 2.34 | 3.00 | 3.66 | 4.35 | 5.13 | 6.06 | 7.29 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 3.07 | 3.83 | 4.57 | 5.35 | 6.21 | 7.23 | 8.56 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 3.82 | 4.67 | 5.49 | 6.35 | 7.28 | 8.38 | 9.80 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 4.59 | 5.53 | 6.42 | 7.34 | 8.35 | 9.52 | 11.0 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 5.38 | 6.39 | 7.36 | 8.34 | 9.41 | 10.7 | 12.2 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 6.18 | 7.27 | 8.30 | 9.34 | 10.5 | 11.8 | 13.4 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 6.99 | 8.15 | 9.24 | 10.3 | 11.5 | 12.9 | 14.6 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 7.81 | 9.03 | 10.2 | 11.3 | 12.6 | 14.0 | 15.8 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 8.63 | 9.93 | 11.1 | 12.3 | 13.6 | 15.1 | 17.0 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 9.47 | 10.8 | 12.1 | 13.3 | 14.7 | 16.2 | 18.2 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 10.3 | 11.7 | 13.0 | 14.3 | 15.7 | 17.3 | 19.3 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 11.2 | 12.6 | 14.0 | 15.3 | 16.8 | 18.4 | 20.5 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 12.0 | 13.5 | 14.9 | 16.3 | 17.8 | 19.5 | 21.6 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 12.9 | 14.4 | 15.9 | 17.3 | 18.9 | 20.6 | 22.8 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 13.7 | 15.4 | 16.9 | 18.3 | 19.9 | 21.7 | 23.9 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 14.6 | 16.3 | 17.8 | 19.3 | 21.0 | 22.8 | 25.0 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 15.4 | 17.2 | 18.8 | 20.3 | 22.0 | 23.9 | 26.2 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 16.3 | 18.1 | 19.7 | 21.3 | 23.0 | 24.9 | 27.3 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 17.2 | 19.0 | 20.7 | 22.3 | 24.1 | 26.0 | 28.4 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 18.1 | 19.9 | 21.7 | 23.3 | 25.1 | 27.1 | 29.6 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 18.9 | 20.9 | 22.6 | 24.3 | 26.1 | 28.2 | 30.7 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 19.8 | 21.8 | 23.6 | 25.3 | 27.2 | 29.2 | 31.8 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 20.7 | 22.7 | 24.5 | 26.3 | 28.2 | 30.3 | 32.9 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 21.6 | 23.6 | 25.5 | 27.3 | 29.2 | 31.4 | 34.0 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 22.5 | 24.6 | 26.5 | 28.3 | 30.3 | 32.5 | 35.1 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 23.4 | 25.5 | 27.4 | 29.3 | 31.3 | 33.5 | 36.3 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |

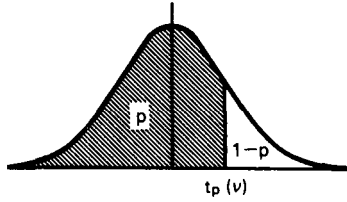
(Table A4 continued on next page.)

TABLE A4 (Cont'd)

| v | p | | | | | | | | | | | | | | | | |
|-----|-------|------|-------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
| 31 | 14.5 | 15.7 | 17.5 | 19.3 | 21.4 | 24.3 | 26.4 | 28.4 | 30.3 | 32.3 | 34.6 | 37.4 | 41.4 | 45.0 | 48.2 | 52.2 | 55.0 |
| 32 | 15.1 | 16.4 | 18.3 | 20.1 | 22.3 | 25.1 | 27.4 | 29.4 | 31.3 | 33.4 | 35.7 | 38.5 | 42.6 | 46.2 | 49.5 | 53.5 | 56.3 |
| 33 | 15.8 | 17.1 | 19.0 | 20.9 | 23.1 | 26.0 | 28.3 | 30.3 | 32.3 | 34.4 | 36.7 | 39.6 | 43.7 | 47.4 | 50.7 | 54.8 | 57.6 |
| 34 | 16.5 | 17.8 | 19.8 | 21.7 | 24.0 | 26.9 | 29.2 | 31.3 | 33.3 | 35.4 | 37.8 | 40.7 | 44.9 | 48.6 | 52.0 | 56.1 | 59.0 |
| 35 | 17.2 | 18.5 | 20.6 | 22.5 | 24.8 | 27.8 | 30.2 | 32.3 | 34.3 | 36.5 | 38.9 | 41.8 | 46.1 | 49.8 | 53.2 | 57.3 | 60.3 |
| 36 | 17.9 | 19.2 | 21.3 | 23.3 | 25.6 | 28.7 | 31.1 | 33.3 | 35.3 | 37.5 | 39.9 | 42.9 | 47.2 | 51.0 | 54.4 | 58.6 | 61.6 |
| 37 | 18.6 | 20.0 | 22.1 | 24.1 | 26.5 | 29.6 | 32.1 | 34.2 | 36.3 | 38.5 | 41.0 | 44.0 | 48.4 | 52.2 | 55.7 | 59.9 | 62.9 |
| 38 | 19.3 | 20.7 | 22.9 | 24.9 | 27.3 | 30.5 | 33.0 | 35.2 | 37.3 | 39.6 | 42.0 | 45.1 | 49.5 | 53.4 | 56.9 | 61.2 | 64.2 |
| 39 | 20.0 | 21.4 | 23.7 | 25.7 | 28.2 | 31.4 | 33.9 | 36.2 | 38.3 | 40.6 | 43.1 | 46.2 | 50.7 | 54.6 | 58.1 | 62.4 | 65.5 |
| 40 | 20.7 | 22.2 | 24.4 | 26.5 | 29.1 | 32.3 | 34.9 | 37.1 | 39.3 | 41.6 | 44.2 | 47.3 | 51.8 | 55.8 | 59.3 | 63.7 | 66.8 |
| 41 | 21.4 | 22.9 | 25.2 | 27.3 | 29.9 | 33.3 | 35.8 | 38.1 | 40.3 | 42.7 | 45.2 | 48.4 | 52.9 | 56.9 | 60.6 | 65.0 | 68.1 |
| 42 | 22.1 | 23.7 | 26.0 | 28.1 | 30.8 | 34.2 | 36.8 | 39.1 | 41.3 | 43.7 | 46.3 | 49.5 | 54.1 | 58.1 | 61.8 | 66.2 | 69.3 |
| 43 | 22.9 | 24.4 | 26.8 | 29.0 | 31.6 | 35.1 | 37.7 | 40.0 | 42.3 | 44.7 | 47.3 | 50.5 | 55.2 | 59.3 | 63.0 | 67.5 | 70.6 |
| 44 | 23.6 | 25.1 | 27.6 | 29.8 | 32.5 | 36.0 | 38.6 | 41.0 | 43.3 | 45.7 | 48.4 | 51.6 | 56.4 | 60.5 | 64.2 | 68.7 | 71.9 |
| 45 | 24.3 | 25.9 | 28.4 | 30.6 | 33.4 | 36.9 | 39.6 | 42.0 | 44.3 | 46.8 | 49.5 | 52.7 | 57.5 | 61.7 | 65.4 | 70.0 | 73.2 |
| 46 | 25.0 | 26.7 | 29.2 | 31.4 | 34.2 | 37.8 | 40.5 | 43.0 | 45.3 | 47.8 | 50.5 | 53.8 | 58.6 | 62.8 | 66.6 | 71.2 | 74.4 |
| 47 | 25.8 | 27.4 | 30.0 | 32.3 | 35.1 | 38.7 | 41.5 | 43.9 | 46.3 | 48.8 | 51.6 | 54.9 | 59.8 | 64.0 | 67.8 | 72.4 | 75.7 |
| 48 | 26.5 | 28.2 | 30.8 | 33.1 | 35.9 | 39.6 | 42.4 | 44.9 | 47.3 | 49.8 | 52.6 | 56.0 | 60.9 | 65.2 | 69.0 | 73.7 | 77.0 |
| 49 | 27.2 | 28.9 | 31.6 | 33.9 | 36.8 | 40.5 | 43.4 | 45.9 | 48.3 | 50.9 | 53.7 | 57.1 | 62.0 | 66.3 | 70.2 | 74.9 | 78.2 |
| 50 | 28.0 | 29.7 | 32.4 | 34.8 | 37.7 | 41.4 | 44.3 | 46.9 | 49.3 | 51.9 | 54.7 | 58.2 | 63.2 | 67.5 | 71.4 | 76.2 | 79.5 |
| 60 | 35.5 | 37.5 | 40.5 | 43.2 | 46.5 | 50.6 | 53.8 | 56.6 | 59.3 | 62.1 | 65.2 | 69.0 | 74.4 | 79.1 | 83.3 | 88.4 | 92.0 |
| 70 | 43.3 | 45.4 | 48.8 | 51.7 | 55.3 | 59.9 | 63.3 | 66.4 | 69.3 | 72.4 | 75.7 | 79.7 | 85.5 | 90.5 | 95.0 | 100.4 | 104.2 |
| 80 | 51.2 | 53.5 | 57.2 | 60.4 | 64.3 | 69.2 | 72.9 | 76.2 | 79.3 | 82.6 | 86.1 | 90.4 | 96.6 | 101.9 | 106.6 | 112.3 | 116.3 |
| 90 | 59.2 | 61.8 | 65.6 | 69.1 | 73.3 | 78.6 | 82.5 | 86.0 | 89.3 | 92.8 | 96.5 | 101.1 | 107.6 | 113.1 | 118.1 | 124.1 | 128.3 |
| 100 | 67.3 | 70.1 | 74.2 | 77.9 | 82.4 | 87.9 | 92.1 | 95.8 | 99.3 | 102.9 | 106.9 | 111.7 | 118.5 | 124.3 | 129.6 | 135.8 | 140.2 |

*Ostle, Bernard, 1975, *Statistics in Research*, 3rd Ed., © 1975, by the Iowa State University Press, Ames, Iowa. The table has been adapted from "A Table of Percentage Points of the χ^2 Distribution," by A. Hald and S. A. Sinkbaek, in *Skandinavisk Aktuarietidskift*, 1950, pp. 170-175. Reprinted with permission of the Iowa State University Press, Ames, Iowa, Almqvist and Wiksell, Int., Stockholm, Sweden (copyright holder for the *Skandinavisk Aktuarietidskift*), and authors Bernard Ostle and Anders Hald.

TABLE A5
Quantile Values $t_p(\nu)$ for the Student's t Distribution
with ν Degrees of Freedom*

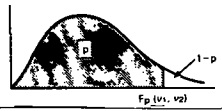


| ν | $t_{0.55}$ | $t_{0.60}$ | $t_{0.70}$ | $t_{0.75}$ | $t_{0.80}$ | $t_{0.90}$ | $t_{0.95}$ | $t_{0.975}$ | $t_{0.99}$ | $t_{0.995}$ |
|----------|------------|------------|------------|------------|------------|------------|------------|-------------|------------|-------------|
| 1 | 0.158 | 0.325 | 0.727 | 1.000 | 1.376 | 3.08 | 6.31 | 12.71 | 31.82 | 63.66 |
| 2 | 0.142 | 0.289 | 0.617 | 0.816 | 1.061 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 |
| 3 | 0.137 | 0.277 | 0.584 | 0.765 | 0.978 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 0.134 | 0.271 | 0.569 | 0.741 | 0.941 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 |
| 5 | 0.132 | 0.267 | 0.559 | 0.727 | 0.920 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 |
| 6 | 0.131 | 0.265 | 0.553 | 0.718 | 0.906 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 |
| 7 | 0.130 | 0.263 | 0.549 | 0.711 | 0.896 | 1.42 | 1.90 | 2.36 | 3.00 | 3.50 |
| 8 | 0.130 | 0.262 | 0.546 | 0.706 | 0.889 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 |
| 9 | 0.129 | 0.261 | 0.543 | 0.703 | 0.883 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 |
| 10 | 0.129 | 0.260 | 0.542 | 0.700 | 0.879 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 |
| 11 | 0.129 | 0.260 | 0.540 | 0.697 | 0.876 | 1.36 | 1.80 | 2.20 | 2.72 | 3.11 |
| 12 | 0.128 | 0.259 | 0.539 | 0.695 | 0.873 | 1.36 | 1.78 | 2.18 | 2.68 | 3.06 |
| 13 | 0.128 | 0.259 | 0.538 | 0.694 | 0.870 | 1.35 | 1.77 | 2.16 | 2.65 | 3.01 |
| 14 | 0.128 | 0.258 | 0.537 | 0.692 | 0.868 | 1.34 | 1.76 | 2.14 | 2.62 | 2.98 |
| 15 | 0.128 | 0.258 | 0.536 | 0.691 | 0.866 | 1.34 | 1.75 | 2.13 | 2.60 | 2.95 |
| 16 | 0.128 | 0.258 | 0.535 | 0.690 | 0.865 | 1.34 | 1.75 | 2.12 | 2.58 | 2.92 |
| 17 | 0.128 | 0.257 | 0.534 | 0.689 | 0.863 | 1.33 | 1.74 | 2.11 | 2.57 | 2.90 |
| 18 | 0.127 | 0.257 | 0.534 | 0.688 | 0.862 | 1.33 | 1.73 | 2.10 | 2.55 | 2.88 |
| 19 | 0.127 | 0.257 | 0.533 | 0.688 | 0.861 | 1.33 | 1.73 | 2.09 | 2.54 | 2.86 |
| 20 | 0.127 | 0.257 | 0.533 | 0.687 | 0.860 | 1.32 | 1.72 | 2.09 | 2.53 | 2.84 |
| 21 | 0.127 | 0.257 | 0.532 | 0.686 | 0.859 | 1.32 | 1.72 | 2.08 | 2.52 | 2.83 |
| 22 | 0.127 | 0.256 | 0.532 | 0.686 | 0.858 | 1.32 | 1.72 | 2.07 | 2.51 | 2.82 |
| 23 | 0.127 | 0.256 | 0.532 | 0.685 | 0.858 | 1.32 | 1.71 | 2.07 | 2.50 | 2.81 |
| 24 | 0.127 | 0.256 | 0.531 | 0.685 | 0.857 | 1.32 | 1.71 | 2.06 | 2.49 | 2.80 |
| 25 | 0.127 | 0.256 | 0.531 | 0.684 | 0.856 | 1.32 | 1.71 | 2.06 | 2.48 | 2.79 |
| 26 | 0.127 | 0.256 | 0.531 | 0.684 | 0.856 | 1.32 | 1.71 | 2.06 | 2.48 | 2.78 |
| 27 | 0.127 | 0.256 | 0.531 | 0.684 | 0.855 | 1.31 | 1.70 | 2.05 | 2.47 | 2.77 |
| 28 | 0.127 | 0.256 | 0.530 | 0.683 | 0.855 | 1.31 | 1.70 | 2.05 | 2.47 | 2.76 |
| 29 | 0.127 | 0.256 | 0.530 | 0.683 | 0.854 | 1.31 | 1.70 | 2.04 | 2.46 | 2.76 |
| 30 | 0.127 | 0.256 | 0.530 | 0.683 | 0.854 | 1.31 | 1.70 | 2.04 | 2.46 | 2.75 |
| 40 | 0.126 | 0.255 | 0.529 | 0.681 | 0.851 | 1.30 | 1.68 | 2.02 | 2.42 | 2.70 |
| 60 | 0.126 | 0.254 | 0.527 | 0.679 | 0.848 | 1.30 | 1.67 | 2.00 | 2.39 | 2.66 |
| 120 | 0.126 | 0.254 | 0.526 | 0.677 | 0.845 | 1.29 | 1.66 | 1.98 | 2.36 | 2.62 |
| ∞ | 0.126 | 0.253 | 0.524 | 0.674 | 0.842 | 1.28 | 1.645 | 1.96 | 2.33 | 2.58 |

*Compiled from Table III of R. A. Fisher and F. Yates, 1974, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th Ed., published by Longman Group UK, Ltd., London (previously published by Oliver and Boyd Ltd., Edinburgh). Reprinted by permission of the publishers.

TABLE A6

Quantile Values $F_p(\nu_1, \nu_2)$ for the F Distribution with ν_1 Numerator Degrees of Freedom and ν_2 Denominator Degrees of Freedom*†



| $\nu_2 \backslash \nu_1$ | ν_1 | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 24 | 30 | 40 | 50 | 60 | 100 | 120 | 200 | 500 | ∞ | |
| 1 | 0.50 | 1.00 | 1.50 | 1.71 | 1.82 | 1.89 | 1.94 | 1.98 | 2.00 | 2.02 | 2.04 | 2.05 | 2.07 | 2.09 | 2.12 | 2.13 | 2.15 | 2.16 | 2.17 | 2.17 | 2.18 | 2.18 | 2.19 | 2.19 | 2.20 |
| | 0.75 | 5.83 | 7.50 | 8.20 | 8.58 | 8.82 | 8.98 | 9.10 | 9.19 | 9.26 | 9.32 | 9.36 | 9.41 | 9.49 | 9.58 | 9.63 | 9.67 | 9.71 | 9.74 | 9.76 | 9.78 | 9.80 | 9.82 | 9.84 | 9.85 |
| | 0.90 | 39.9 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 | 58.9 | 59.4 | 59.9 | 60.2 | 60.5 | 60.7 | 61.2 | 61.7 | 62.0 | 62.3 | 62.5 | 62.7 | 62.8 | 63.0 | 63.1 | 63.2 | 63.3 | 63.3 |
| | 0.95 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 243 | 244 | 246 | 248 | 249 | 250 | 252 | 252 | 252 | 253 | 253 | 254 | 254 | 254 |
| | 0.975 | 648 | 800 | 864 | 900 | 922 | 937 | 948 | 957 | 963 | 966 | 971 | 974 | 985 | 993 | 997 | 1000 | 1010 | 1010 | 1010 | 1010 | 1010 | 1020 | 1020 | 1020 |
| 0.99 | 4050 | 5000 | 5400 | 5620 | 5760 | 5860 | 5930 | 5980 | 6020 | 6060 | 6080 | 6110 | 6160 | 6210 | 6230 | 6260 | 6290 | 6300 | 6310 | 6330 | 6340 | 6350 | 6360 | 6370 | |
| 0.995 | 6200 | 20000 | 21600 | 22500 | 23100 | 23400 | 23700 | 23900 | 24100 | 24200 | 24300 | 24400 | 24600 | 24800 | 24900 | 25000 | 25100 | 25200 | 25300 | 25400 | 25400 | 25400 | 25400 | 25400 | |
| 2 | 0.50 | 0.667 | 1.00 | 1.13 | 1.21 | 1.25 | 1.28 | 1.30 | 1.32 | 1.33 | 1.34 | 1.35 | 1.36 | 1.38 | 1.39 | 1.40 | 1.41 | 1.42 | 1.42 | 1.43 | 1.43 | 1.43 | 1.44 | 1.44 | 1.44 |
| | 0.75 | 2.57 | 3.00 | 3.15 | 3.23 | 3.28 | 3.31 | 3.34 | 3.35 | 3.37 | 3.38 | 3.40 | 3.41 | 3.43 | 3.43 | 3.44 | 3.45 | 3.45 | 3.45 | 3.46 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 |
| | 0.90 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 | 9.39 | 9.40 | 9.41 | 9.42 | 9.44 | 9.45 | 9.46 | 9.47 | 9.47 | 9.48 | 9.48 | 9.49 | 9.49 | 9.49 | 9.49 |
| | 0.95 | 18.5 | 19.0 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 |
| | 0.975 | 38.5 | 39.0 | 39.2 | 39.2 | 39.3 | 39.3 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 |
| 0.99 | 98.5 | 99.0 | 99.2 | 99.2 | 99.3 | 99.3 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | |
| 0.995 | 198 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 200 | |
| 3 | 0.50 | 0.585 | 0.891 | 1.00 | 1.06 | 1.10 | 1.13 | 1.15 | 1.16 | 1.17 | 1.18 | 1.19 | 1.20 | 1.21 | 1.23 | 1.23 | 1.24 | 1.25 | 1.25 | 1.25 | 1.26 | 1.26 | 1.27 | 1.27 | |
| | 0.75 | 2.02 | 2.28 | 2.36 | 2.39 | 2.41 | 2.42 | 2.43 | 2.44 | 2.44 | 2.44 | 2.45 | 2.45 | 2.46 | 2.46 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | |
| | 0.90 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 | 5.23 | 5.22 | 5.22 | 5.20 | 5.18 | 5.18 | 5.17 | 5.16 | 5.15 | 5.15 | 5.14 | 5.14 | 5.14 | 5.13 | |
| | 0.95 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.76 | 8.74 | 8.70 | 8.66 | 8.63 | 8.62 | 8.59 | 8.58 | 8.57 | 8.55 | 8.55 | 8.54 | 8.53 | 8.53 |
| | 0.975 | 17.4 | 16.0 | 15.4 | 15.1 | 14.9 | 14.7 | 14.6 | 14.5 | 14.5 | 14.4 | 14.4 | 14.3 | 14.3 | 14.2 | 14.1 | 14.1 | 14.0 | 14.0 | 14.0 | 13.9 | 13.9 | 13.9 | 13.9 | |
| 0.99 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 | 27.5 | 27.3 | 27.2 | 27.1 | 27.1 | 26.9 | 26.7 | 26.6 | 26.5 | 26.4 | 26.4 | 26.3 | 26.2 | 26.2 | 26.2 | 26.1 | 26.1 | |
| 0.995 | 55.6 | 49.8 | 47.5 | 46.2 | 45.4 | 44.8 | 44.4 | 44.1 | 43.9 | 43.7 | 43.5 | 43.4 | 43.1 | 42.8 | 42.6 | 42.5 | 42.3 | 42.2 | 42.1 | 42.0 | 41.9 | 41.9 | 41.8 | | |
| 4 | 0.50 | 0.549 | 0.828 | 0.941 | 1.00 | 1.04 | 1.06 | 1.08 | 1.09 | 1.10 | 1.11 | 1.12 | 1.13 | 1.14 | 1.15 | 1.16 | 1.16 | 1.17 | 1.18 | 1.18 | 1.18 | 1.18 | 1.19 | 1.19 | |
| | 0.75 | 1.81 | 2.00 | 2.05 | 2.06 | 2.07 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | |
| | 0.90 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 | 3.92 | 3.91 | 3.90 | 3.87 | 3.84 | 3.83 | 3.82 | 3.80 | 3.80 | 3.79 | 3.78 | 3.78 | 3.77 | 3.76 | 3.76 |
| | 0.95 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.94 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.70 | 5.69 | 5.66 | 5.66 | 5.65 | 5.64 | 5.63 |
| | 0.975 | 12.2 | 10.6 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.90 | 8.84 | 8.79 | 8.75 | 8.66 | 8.56 | 8.51 | 8.46 | 8.41 | 8.38 | 8.36 | 8.32 | 8.31 | 8.29 | 8.27 | 8.26 |
| 0.99 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 | 14.8 | 14.7 | 14.5 | 14.4 | 14.4 | 14.2 | 14.0 | 13.9 | 13.8 | 13.7 | 13.7 | 13.7 | 13.6 | 13.6 | 13.5 | 13.5 | 13.5 | |
| 0.995 | 31.3 | 26.3 | 24.3 | 23.2 | 22.5 | 22.0 | 21.6 | 21.4 | 21.1 | 21.0 | 20.8 | 20.7 | 20.4 | 20.2 | 20.0 | 19.9 | 19.8 | 19.7 | 19.6 | 19.5 | 19.5 | 19.4 | 19.4 | 19.3 | |
| 5 | 0.50 | 0.528 | 0.799 | 0.907 | 0.965 | 1.00 | 1.02 | 1.04 | 1.05 | 1.06 | 1.07 | 1.08 | 1.09 | 1.10 | 1.11 | 1.12 | 1.12 | 1.13 | 1.13 | 1.14 | 1.14 | 1.14 | 1.15 | 1.15 | |
| | 0.75 | 1.69 | 1.85 | 1.88 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.88 | 1.88 | 1.88 | 1.88 | 1.88 | 1.87 | 1.87 | 1.87 | 1.87 | 1.87 | |
| | 0.90 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 | 3.30 | 3.29 | 3.27 | 3.24 | 3.21 | 3.19 | 3.17 | 3.16 | 3.15 | 3.14 | 3.13 | 3.12 | 3.12 | 3.11 | 3.10 |
| | 0.95 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.71 | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.44 | 4.43 | 4.41 | 4.40 | 4.39 | 4.37 | 4.36 |
| | 0.975 | 10.10 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.68 | 6.62 | 6.57 | 6.52 | 6.43 | 6.33 | 6.28 | 6.23 | 6.18 | 6.14 | 6.12 | 6.08 | 6.07 | 6.05 | 6.03 | 6.02 |
| 0.99 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 | 10.3 | 10.2 | 10.1 | 9.96 | 9.89 | 9.72 | 9.55 | 9.47 | 9.38 | 9.29 | 9.24 | 9.20 | 9.13 | 9.11 | 9.08 | 9.04 | 9.02 | |
| 0.995 | 22.8 | 18.3 | 16.5 | 15.6 | 14.9 | 14.5 | 14.2 | 14.0 | 13.8 | 13.6 | 13.5 | 13.4 | 13.1 | 12.9 | 12.8 | 12.7 | 12.5 | 12.5 | 12.4 | 12.3 | 12.2 | 12.2 | 12.1 | | |
| 6 | 0.50 | 0.515 | 0.780 | 0.88 | 0.942 | 0.977 | 1.00 | 1.02 | 1.03 | 1.04 | 1.05 | 1.05 | 1.06 | 1.07 | 1.08 | 1.09 | 1.10 | 1.10 | 1.11 | 1.11 | 1.11 | 1.12 | 1.12 | 1.12 | |
| | 0.75 | 1.62 | 1.76 | 1.78 | 1.79 | 1.79 | 1.78 | 1.78 | 1.78 | 1.77 | 1.77 | 1.77 | 1.77 | 1.76 | 1.76 | 1.75 | 1.75 | 1.75 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | |
| | 0.90 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 | 2.94 | 2.92 | 2.90 | 2.87 | 2.84 | 2.82 | 2.80 | 2.78 | 2.77 | 2.76 | 2.75 | 2.74 | 2.73 | 2.73 | 2.72 |
| | 0.95 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.03 | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.75 | 3.74 | 3.71 | 3.70 | 3.69 | 3.68 | 3.67 |
| | 0.975 | 8.81 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.52 | 5.46 | 5.41 | 5.37 | 5.27 | 5.17 | 5.12 | 5.07 | 5.01 | 4.98 | 4.96 | 4.92 | 4.90 | 4.88 | 4.86 | 4.85 |
| 0.99 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.79 | 7.72 | 7.56 | 7.40 | 7.31 | 7.23 | 7.14 | 7.09 | 7.06 | 6.99 | 6.97 | 6.93 | 6.90 | 6.88 | |
| 0.995 | 18.6 | 14.5 | 12.9 | 12.0 | 11.5 | 11.1 | 10.8 | 10.6 | 10.4 | 10.2 | 10.1 | 10.0 | 9.81 | 9.59 | 9.47 | 9.36 | 9.24 | 9.17 | 9.12 | 9.03 | 9.00 | 8.95 | 8.91 | 8.88 | |

| v ₂ | p | N ₁ | | | | | | | | | | | | | | | | | | | | | | | | |
|----------------|-------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 24 | 30 | 40 | 50 | 60 | 100 | 120 | 200 | 500 | ∞ | |
| 7 | 0.50 | 0.506 | 0.767 | 0.871 | 0.926 | 0.960 | 0.983 | 1.00 | 1.01 | 1.02 | 1.03 | 1.04 | 1.04 | 1.05 | 1.07 | 1.07 | 1.08 | 1.08 | 1.09 | 1.09 | 1.10 | 1.10 | 1.10 | 1.10 | 1.10 | 1.10 |
| | 0.75 | 1.57 | 1.70 | 1.72 | 1.72 | 1.71 | 1.71 | 1.70 | 1.70 | 1.69 | 1.69 | 1.69 | 1.68 | 1.68 | 1.67 | 1.67 | 1.66 | 1.66 | 1.66 | 1.65 | 1.65 | 1.65 | 1.65 | 1.65 | 1.65 | 1.65 |
| | 0.90 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 | 2.70 | 2.68 | 2.67 | 2.63 | 2.59 | 2.58 | 2.56 | 2.54 | 2.52 | 2.51 | 2.50 | 2.49 | 2.48 | 2.48 | 2.47 | 2.47 |
| | 0.95 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.60 | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.32 | 3.30 | 3.27 | 3.27 | 3.25 | 3.24 | 3.24 | 3.23 |
| | 0.975 | 8.07 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.82 | 4.76 | 4.71 | 4.67 | 4.57 | 4.47 | 4.42 | 4.36 | 4.31 | 4.28 | 4.25 | 4.21 | 4.20 | 4.18 | 4.16 | 4.14 | 4.14 |
| 0.99 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.88 | 6.72 | 6.62 | 6.54 | 6.47 | 6.31 | 6.16 | 6.07 | 5.99 | 5.94 | 5.86 | 5.79 | 5.74 | 5.69 | 5.67 | 5.64 | 5.61 | 5.61 | |
| 0.995 | 16.2 | 12.4 | 10.9 | 10.0 | 9.52 | 9.16 | 8.89 | 8.68 | 8.51 | 8.38 | 8.27 | 8.18 | 7.97 | 7.75 | 7.65 | 7.53 | 7.42 | 7.35 | 7.31 | 7.22 | 7.19 | 7.15 | 7.10 | 7.10 | 7.08 | |
| 8 | 0.50 | 0.499 | 0.757 | 0.860 | 0.915 | 0.948 | 0.971 | 0.988 | 1.00 | 1.01 | 1.02 | 1.02 | 1.03 | 1.04 | 1.05 | 1.06 | 1.07 | 1.07 | 1.07 | 1.08 | 1.08 | 1.08 | 1.09 | 1.09 | 1.09 | |
| | 0.75 | 1.54 | 1.66 | 1.67 | 1.66 | 1.66 | 1.65 | 1.64 | 1.64 | 1.64 | 1.63 | 1.63 | 1.62 | 1.62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.58 | 1.58 | 1.58 | 1.58 | 1.58 | |
| | 0.90 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 | 2.54 | 2.52 | 2.50 | 2.46 | 2.42 | 2.40 | 2.38 | 2.36 | 2.35 | 2.34 | 2.32 | 2.32 | 2.31 | 2.30 | 2.29 | |
| | 0.95 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.31 | 3.28 | 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 3.02 | 3.01 | 2.97 | 2.97 | 2.95 | 2.94 | 2.93 | |
| | 0.975 | 7.57 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.36 | 4.30 | 4.24 | 4.20 | 4.10 | 4.00 | 3.95 | 3.89 | 3.84 | 3.81 | 3.78 | 3.74 | 3.73 | 3.70 | 3.68 | 3.67 | |
| 0.99 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.73 | 5.67 | 5.52 | 5.36 | 5.28 | 5.20 | 5.12 | 5.07 | 5.03 | 4.96 | 4.95 | 4.91 | 4.88 | 4.86 | | |
| 0.995 | 14.7 | 11.0 | 9.60 | 8.81 | 8.30 | 7.95 | 7.69 | 7.50 | 7.34 | 7.21 | 7.10 | 7.01 | 6.81 | 6.61 | 6.50 | 6.40 | 6.29 | 6.22 | 6.18 | 6.09 | 6.06 | 6.02 | 5.98 | 5.95 | | |
| 9 | 0.50 | 0.494 | 0.749 | 0.852 | 0.906 | 0.939 | 0.962 | 0.978 | 0.990 | 1.00 | 1.01 | 1.01 | 1.02 | 1.03 | 1.04 | 1.05 | 1.05 | 1.06 | 1.06 | 1.07 | 1.07 | 1.07 | 1.08 | 1.08 | 1.08 | |
| | 0.75 | 1.51 | 1.62 | 1.63 | 1.63 | 1.62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.58 | 1.58 | 1.58 | 1.57 | 1.56 | 1.56 | 1.55 | 1.55 | 1.54 | 1.54 | 1.53 | 1.53 | 1.53 | 1.53 | 1.53 | |
| | 0.90 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 | 2.42 | 2.40 | 2.38 | 2.34 | 2.30 | 2.28 | 2.25 | 2.23 | 2.22 | 2.21 | 2.19 | 2.18 | 2.17 | 2.17 | 2.16 | |
| | 0.95 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.10 | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2.80 | 2.79 | 2.76 | 2.75 | 2.73 | 2.72 | 2.71 | |
| | 0.975 | 7.21 | 5.71 | 5.08 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 4.03 | 3.96 | 3.91 | 3.87 | 3.77 | 3.67 | 3.61 | 3.56 | 3.51 | 3.47 | 3.45 | 3.40 | 3.39 | 3.37 | 3.35 | 3.33 | |
| 0.99 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.18 | 5.11 | 4.96 | 4.81 | 4.73 | 4.65 | 4.57 | 4.52 | 4.48 | 4.42 | 4.40 | 4.36 | 4.33 | 4.31 | | |
| 0.995 | 13.6 | 10.1 | 8.72 | 7.96 | 7.47 | 7.13 | 6.88 | 6.69 | 6.54 | 6.42 | 6.31 | 6.23 | 6.03 | 5.83 | 5.73 | 5.62 | 5.52 | 5.45 | 5.41 | 5.32 | 5.30 | 5.26 | 5.21 | 5.19 | | |
| 10 | 0.490 | 0.743 | 0.845 | 0.899 | 0.932 | 0.954 | 0.971 | 0.983 | 0.992 | 1.00 | 1.01 | 1.01 | 1.02 | 1.02 | 1.03 | 1.04 | 1.05 | 1.05 | 1.06 | 1.06 | 1.06 | 1.07 | 1.07 | 1.07 | 1.07 | |
| | 0.75 | 1.49 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.57 | 1.56 | 1.55 | 1.55 | 1.55 | 1.54 | 1.53 | 1.52 | 1.52 | 1.51 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 | 1.47 | |
| | 0.90 | 3.28 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 | 2.30 | 2.28 | 2.24 | 2.20 | 2.18 | 2.16 | 2.13 | 2.12 | 2.11 | 2.09 | 2.08 | 2.07 | 2.06 | 2.06 | |
| | 0.95 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.94 | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.64 | 2.62 | 2.59 | 2.58 | 2.56 | 2.55 | 2.54 | |
| | 0.975 | 6.94 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.78 | 3.72 | 3.66 | 3.62 | 3.52 | 3.42 | 3.37 | 3.31 | 3.26 | 3.22 | 3.20 | 3.15 | 3.14 | 3.17 | 3.09 | 3.08 | |
| 0.99 | 10.0 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.77 | 4.71 | 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.12 | 4.08 | 4.01 | 4.00 | 3.96 | 3.92 | 3.91 | | |
| 0.995 | 12.8 | 9.43 | 8.08 | 7.34 | 6.87 | 6.54 | 6.30 | 6.12 | 5.97 | 5.85 | 5.75 | 5.66 | 5.47 | 5.27 | 5.17 | 5.07 | 4.97 | 4.90 | 4.86 | 4.77 | 4.75 | 4.71 | 4.67 | 4.64 | | |
| 11 | 0.486 | 0.739 | 0.840 | 0.893 | 0.926 | 0.948 | 0.964 | 0.977 | 0.986 | 0.994 | 1.00 | 1.01 | 1.02 | 1.03 | 1.03 | 1.04 | 1.05 | 1.05 | 1.05 | 1.06 | 1.06 | 1.06 | 1.06 | 1.06 | 1.06 | |
| | 0.75 | 1.47 | 1.58 | 1.58 | 1.57 | 1.56 | 1.55 | 1.54 | 1.53 | 1.53 | 1.52 | 1.52 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 | 1.47 | 1.46 | 1.46 | 1.46 | 1.45 | 1.45 | |
| | 0.90 | 3.23 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 | 2.23 | 2.21 | 2.17 | 2.12 | 2.10 | 2.08 | 2.05 | 2.04 | 2.03 | 2.00 | 2.00 | 1.99 | 1.98 | 1.97 | |
| | 0.95 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.82 | 2.79 | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.51 | 2.49 | 2.46 | 2.45 | 2.43 | 2.42 | 2.40 | |
| | 0.975 | 6.72 | 5.26 | 4.63 | 4.28 | 4.04 | 3.88 | 3.76 | 3.66 | 3.59 | 3.53 | 3.47 | 3.43 | 3.33 | 3.23 | 3.17 | 3.12 | 3.06 | 3.03 | 3.00 | 2.96 | 2.94 | 2.92 | 2.90 | 2.88 | |
| 0.99 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.46 | 4.40 | 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.81 | 3.78 | 3.71 | 3.69 | 3.66 | 3.62 | 3.60 | | |
| 0.995 | 12.2 | 8.91 | 7.60 | 6.88 | 6.42 | 6.10 | 5.86 | 5.68 | 5.54 | 5.42 | 5.32 | 5.24 | 5.05 | 4.86 | 4.76 | 4.65 | 4.55 | 4.49 | 4.45 | 4.36 | 4.34 | 4.29 | 4.25 | 4.23 | | |
| 12 | 0.484 | 0.735 | 0.835 | 0.888 | 0.921 | 0.943 | 0.959 | 0.972 | 0.981 | 0.989 | 0.995 | 1.00 | 1.01 | 1.02 | 1.03 | 1.03 | 1.04 | 1.04 | 1.05 | 1.05 | 1.05 | 1.05 | 1.06 | 1.06 | 1.06 | |
| | 0.75 | 1.46 | 1.56 | 1.56 | 1.55 | 1.54 | 1.53 | 1.52 | 1.51 | 1.51 | 1.50 | 1.50 | 1.49 | 1.48 | 1.47 | 1.46 | 1.45 | 1.45 | 1.44 | 1.44 | 1.43 | 1.43 | 1.43 | 1.42 | 1.42 | |
| | 0.90 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.21 | 2.19 | 2.17 | 2.15 | 2.11 | 2.06 | 2.04 | 2.01 | 1.99 | 1.97 | 1.96 | 1.94 | 1.93 | 1.92 | 1.91 | 1.90 | |
| | 0.95 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.72 | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.40 | 2.38 | 2.35 | 2.34 | 2.32 | 2.31 | 2.30 | |
| | 0.975 | 6.55 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.44 | 3.37 | 3.32 | 3.28 | 3.18 | 3.07 | 3.02 | 2.96 | 2.91 | 2.87 | 2.85 | 2.80 | 2.79 | 2.76 | 2.74 | 2.72 | |
| 0.99 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.22 | 4.16 | 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.57 | 3.54 | 3.47 | 3.45 | 3.41 | 3.38 | 3.36 | | |
| 0.995 | 11.8 | 8.51 | 7.23 | 6.52 | 6.07 | 5.76 | 5.52 | 5.35 | 5.20 | 5.09 | 4.99 | 4.91 | 4.72 | 4.53 | 4.43 | 4.33 | 4.23 | 4.17 | 4.12 | 4.04 | 4.01 | 3.97 | 3.93 | 3.90 | | |
| 15 | 0.478 | 0.726 | 0.826 | 0.878 | 0.911 | 0.933 | 0.948 | 0.960 | 0.970 | 0.977 | 0.984 | 0.989 | 1.00 | 1.01 | 1.02 | 1.02 | 1.03 | 1.03 | 1.03 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 | 1.05 | |
| | 0.75 | 1.43 | 1.52 | 1.52 | 1.51 | 1.49 | 1.48 | 1.47 | 1.46 | 1.46 | 1.45 | 1.44 | 1.44 | 1.43 | 1.41 | 1.41 | 1.40 | 1.39 | 1.39 | 1.38 | 1.38 | 1.37 | 1.37 | 1.36 | 1.36 | |
| | 0.90 | 3.07 | 2.70 | 2.49 | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.09 | 2.06 | 2.04 | 2.02 | 1.97 | 1.92 | 1.90 | 1.87 | 1.85 | 1.83 | 1.82 | 1.79 | 1.79 | 1.77 | 1.76 | 1.76 | |
| | 0.95 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.51 | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.18 | 2.16 | 2.12 | 2.11 | 2.10 | 2.08 | 2.07 | |
| | 0.975 | 6.20 | 4.76 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 | 3.06 | 3.01 | 2.96 | 2.86 | 2.76 | 2.70 | 2.64 | 2.59 | 2.55 | 2.52 | 2.47 | 2.46 | 2.44 | 2.41 | 2.40 | |
| 0.99 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.73 | 3.67 | 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.08 | 3.05 | 2.98 | 2.96 | 2.92 | 2.89 | 2.87 | | |
| 0.995 | 10.3 | 7.70 | 6.48 | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 | 4.42 | 4.33 | 4.25 | 4.07 | 3.88 | 3.79 | 3.69 | 3.59 | 3.52 | 3.48 | 3.39 | 3.37 | 3.33 | 3.29 | 3.26 | | |
| 20 | 0.472 | 0.718 | 0.816 | 0.868 | 0.900 | 0.922 | 0.938 | 0.950 | 0.959 | 0.966 | 0.972 | 0.977 | 0.989 | 1.00 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | 1.03 | |
| | 0.75 | 1.40 | 1.49 | 1.48 | 1.47 | 1.45 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.33 | 1.32 | 1.31 | 1.31 | 1.30 | 1.30 | 1.29 | |

TABLE A6 (Cont'd)

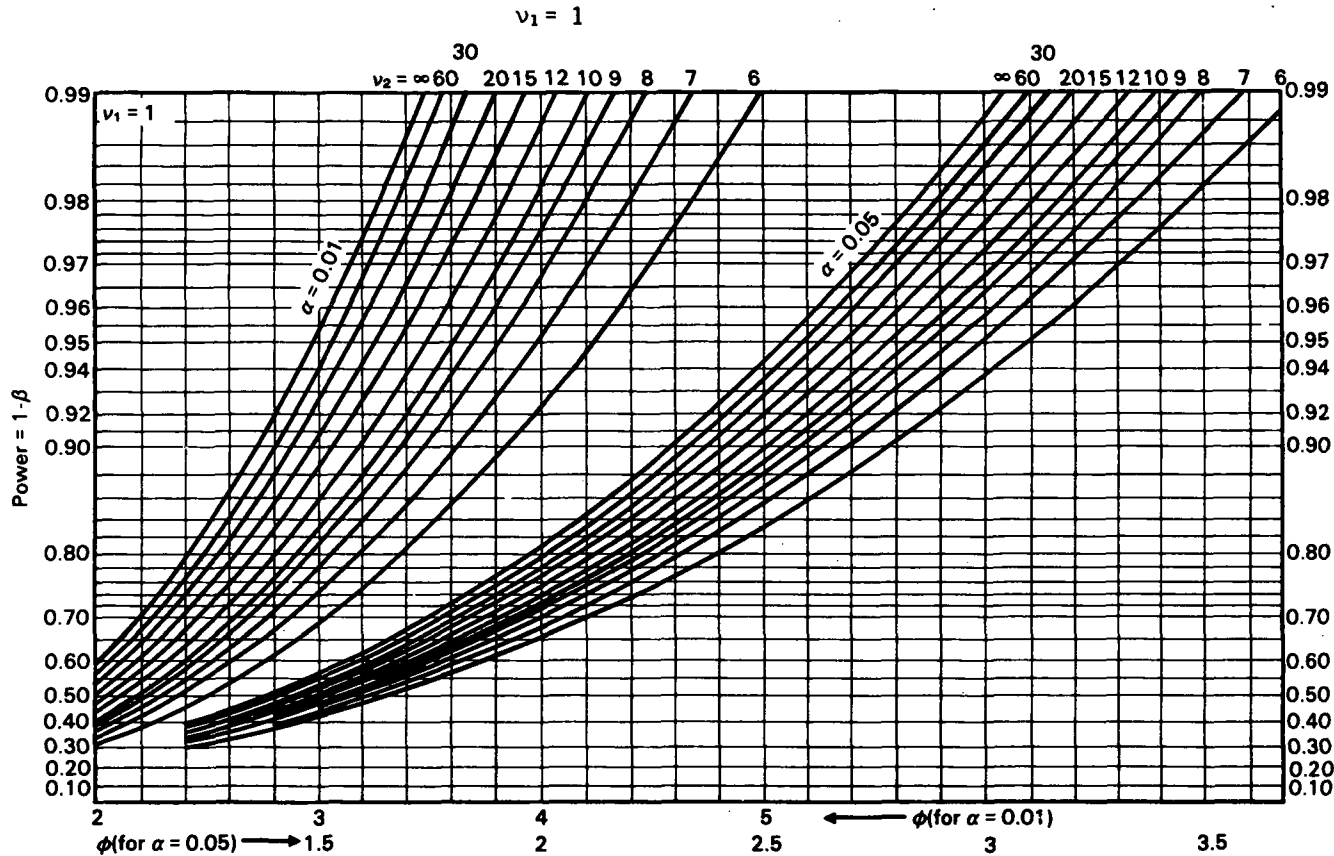
| v_2 | v_1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 24 | 30 | 40 | 50 | 60 | 100 | 120 | 200 | 500 | ∞ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| 24 | 0.50 | 0.469 | 0.714 | 0.812 | 0.863 | 0.895 | 0.917 | 0.932 | 0.944 | 0.953 | 0.961 | 0.967 | 0.972 | 0.983 | 0.994 | 1.00 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.03 | 1.03 |
| | 0.75 | 1.39 | 1.47 | 1.46 | 1.44 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.38 | 1.37 | 1.36 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.29 | 1.28 | 1.28 | 1.27 | 1.27 | 1.26 |
| | 0.90 | 2.93 | 2.54 | 2.33 | 2.19 | 2.10 | 2.04 | 1.98 | 1.94 | 1.91 | 1.88 | 1.85 | 1.83 | 1.78 | 1.73 | 1.70 | 1.67 | 1.64 | 1.62 | 1.61 | 1.58 | 1.57 | 1.56 | 1.54 | 1.53 |
| | 0.95 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.21 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.86 | 1.84 | 1.80 | 1.79 | 1.77 | 1.75 | 1.73 |
| | 0.975 | 5.72 | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 | 2.64 | 2.59 | 2.54 | 2.44 | 2.33 | 2.27 | 2.21 | 2.15 | 2.11 | 2.08 | 2.02 | 2.01 | 1.98 | 1.95 | 1.94 |
| 0.99 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.09 | 3.02 | 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.44 | 2.40 | 2.33 | 2.31 | 2.27 | 2.24 | 2.21 | |
| 0.995 | 9.55 | 6.66 | 5.52 | 4.89 | 4.49 | 4.20 | 3.99 | 3.83 | 3.69 | 3.59 | 3.50 | 3.42 | 3.25 | 3.06 | 2.97 | 2.87 | 2.77 | 2.70 | 2.66 | 2.57 | 2.55 | 2.50 | 2.46 | 2.43 | |
| 30 | 0.50 | 0.466 | 0.709 | 0.807 | 0.858 | 0.890 | 0.912 | 0.927 | 0.939 | 0.948 | 0.955 | 0.961 | 0.966 | 0.978 | 0.989 | 0.994 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 |
| | 0.75 | 1.38 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.35 | 1.34 | 1.32 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.26 | 1.25 | 1.24 | 1.24 | 1.23 | 1.23 |
| | 0.90 | 2.88 | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.93 | 1.88 | 1.85 | 1.82 | 1.79 | 1.77 | 1.72 | 1.67 | 1.64 | 1.61 | 1.57 | 1.55 | 1.54 | 1.51 | 1.50 | 1.48 | 1.47 | 1.46 |
| | 0.95 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.13 | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.76 | 1.74 | 1.70 | 1.68 | 1.66 | 1.64 | 1.62 |
| | 0.975 | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 | 2.46 | 2.41 | 2.31 | 2.20 | 2.14 | 2.07 | 2.01 | 1.97 | 1.94 | 1.88 | 1.87 | 1.84 | 1.81 | 1.79 |
| 0.99 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.91 | 2.84 | 2.70 | 2.55 | 2.47 | 2.39 | 2.30 | 2.25 | 2.21 | 2.13 | 2.11 | 2.07 | 2.03 | 2.01 | |
| 0.995 | 9.18 | 6.35 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 | 3.34 | 3.25 | 3.18 | 3.01 | 2.82 | 2.73 | 2.63 | 2.52 | 2.46 | 2.42 | 2.32 | 2.30 | 2.25 | 2.21 | 2.18 | |
| 40 | 0.50 | 0.463 | 0.705 | 0.802 | 0.854 | 0.885 | 0.907 | 0.922 | 0.934 | 0.943 | 0.950 | 0.956 | 0.961 | 0.972 | 0.983 | 0.989 | 0.994 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.01 | 1.02 | 1.02 |
| | 0.75 | 1.36 | 1.44 | 1.42 | 1.40 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.28 | 1.26 | 1.25 | 1.24 | 1.23 | 1.22 | 1.21 | 1.21 | 1.20 | 1.19 | 1.19 | |
| | 0.90 | 2.84 | 2.44 | 2.23 | 2.09 | 2.00 | 1.93 | 1.87 | 1.83 | 1.79 | 1.76 | 1.73 | 1.71 | 1.66 | 1.61 | 1.57 | 1.54 | 1.51 | 1.48 | 1.47 | 1.43 | 1.42 | 1.41 | 1.39 | 1.38 |
| | 0.95 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.04 | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.66 | 1.64 | 1.59 | 1.58 | 1.55 | 1.53 | 1.51 |
| | 0.975 | 5.42 | 4.05 | 3.46 | 3.13 | 2.90 | 2.74 | 2.62 | 2.53 | 2.45 | 2.39 | 2.33 | 2.29 | 2.18 | 2.07 | 2.01 | 1.94 | 1.88 | 1.83 | 1.80 | 1.74 | 1.72 | 1.69 | 1.66 | 1.64 |
| 0.99 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.73 | 2.66 | 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.06 | 2.02 | 1.94 | 1.92 | 1.87 | 1.83 | 1.80 | |
| 0.995 | 8.83 | 6.07 | 4.98 | 4.37 | 3.99 | 3.71 | 3.51 | 3.35 | 3.22 | 3.12 | 3.03 | 2.95 | 2.78 | 2.60 | 2.50 | 2.40 | 2.30 | 2.23 | 2.18 | 2.09 | 2.06 | 2.01 | 1.96 | 1.93 | |
| 60 | 0.50 | 0.461 | 0.701 | 0.798 | 0.849 | 0.880 | 0.901 | 0.917 | 0.928 | 0.937 | 0.945 | 0.951 | 0.956 | 0.967 | 0.978 | 0.983 | 0.989 | 0.994 | 0.998 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.01 |
| | 0.75 | 1.35 | 1.42 | 1.41 | 1.38 | 1.37 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.27 | 1.25 | 1.24 | 1.22 | 1.21 | 1.20 | 1.19 | 1.17 | 1.17 | 1.16 | 1.15 | 1.15 | |
| | 0.90 | 2.79 | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.82 | 1.77 | 1.74 | 1.71 | 1.68 | 1.66 | 1.60 | 1.54 | 1.51 | 1.48 | 1.44 | 1.41 | 1.40 | 1.36 | 1.35 | 1.33 | 1.31 | 1.29 |
| | 0.95 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.95 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.56 | 1.53 | 1.48 | 1.47 | 1.44 | 1.41 | 1.39 |
| | 0.975 | 5.29 | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 | 2.27 | 2.22 | 2.17 | 2.06 | 1.94 | 1.88 | 1.82 | 1.74 | 1.70 | 1.67 | 1.60 | 1.58 | 1.54 | 1.51 | 1.48 |
| 0.99 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.56 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.88 | 1.84 | 1.75 | 1.73 | 1.68 | 1.63 | 1.60 | |
| 0.995 | 8.49 | 5.80 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 | 2.90 | 2.82 | 2.74 | 2.57 | 2.39 | 2.29 | 2.19 | 2.08 | 2.01 | 1.96 | 1.86 | 1.83 | 1.78 | 1.73 | 1.69 | |
| 120 | 0.50 | 0.458 | 0.697 | 0.793 | 0.844 | 0.875 | 0.896 | 0.912 | 0.923 | 0.932 | 0.939 | 0.945 | 0.950 | 0.961 | 0.972 | 0.978 | 0.983 | 0.989 | 0.992 | 0.994 | 1.00 | 1.00 | 1.00 | 1.01 | 1.01 |
| | 0.75 | 1.34 | 1.40 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.24 | 1.22 | 1.21 | 1.19 | 1.18 | 1.17 | 1.16 | 1.14 | 1.13 | 1.12 | 1.11 | 1.10 |
| | 0.90 | 2.75 | 2.35 | 2.13 | 1.99 | 1.90 | 1.82 | 1.77 | 1.72 | 1.68 | 1.65 | 1.62 | 1.60 | 1.55 | 1.48 | 1.45 | 1.41 | 1.37 | 1.34 | 1.32 | 1.27 | 1.26 | 1.24 | 1.21 | 1.19 |
| | 0.95 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 | 1.87 | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.46 | 1.43 | 1.37 | 1.35 | 1.32 | 1.28 | 1.25 |
| | 0.975 | 5.15 | 3.80 | 3.23 | 2.89 | 2.67 | 2.52 | 2.39 | 2.30 | 2.22 | 2.16 | 2.10 | 2.05 | 1.95 | 1.82 | 1.76 | 1.69 | 1.61 | 1.56 | 1.53 | 1.45 | 1.43 | 1.39 | 1.34 | 1.31 |
| 0.99 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.40 | 2.34 | 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 1.70 | 1.66 | 1.56 | 1.53 | 1.48 | 1.42 | 1.38 | |
| 0.995 | 8.18 | 5.54 | 4.50 | 3.92 | 3.55 | 3.28 | 3.09 | 2.93 | 2.81 | 2.71 | 2.62 | 2.54 | 2.37 | 2.19 | 2.09 | 1.98 | 1.87 | 1.80 | 1.75 | 1.64 | 1.61 | 1.54 | 1.48 | 1.43 | |
| ∞ | 0.50 | 0.455 | 0.693 | 0.789 | 0.839 | 0.870 | 0.891 | 0.907 | 0.918 | 0.927 | 0.934 | 0.939 | 0.945 | 0.956 | 0.967 | 0.972 | 0.978 | 0.983 | 0.987 | 0.989 | 0.993 | 0.994 | 0.997 | 0.999 | 1.00 |
| | 0.75 | 1.32 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.29 | 1.28 | 1.27 | 1.25 | 1.24 | 1.22 | 1.19 | 1.18 | 1.16 | 1.14 | 1.13 | 1.12 | 1.12 | 1.09 | 1.08 | 1.07 | 1.04 | 1.00 |
| | 0.90 | 2.71 | 2.30 | 2.08 | 1.94 | 1.85 | 1.77 | 1.72 | 1.67 | 1.63 | 1.60 | 1.57 | 1.55 | 1.49 | 1.42 | 1.38 | 1.34 | 1.30 | 1.26 | 1.24 | 1.18 | 1.17 | 1.13 | 1.08 | 1.00 |
| | 0.95 | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.79 | 1.75 | 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.35 | 1.32 | 1.24 | 1.22 | 1.17 | 1.11 | 1.00 |
| | 0.975 | 5.02 | 3.69 | 3.12 | 2.79 | 2.57 | 2.41 | 2.29 | 2.19 | 2.11 | 2.05 | 1.99 | 1.94 | 1.83 | 1.71 | 1.64 | 1.57 | 1.48 | 1.43 | 1.39 | 1.30 | 1.27 | 1.21 | 1.13 | 1.00 |
| 0.99 | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.25 | 2.18 | 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.52 | 1.47 | 1.36 | 1.32 | 1.25 | 1.15 | 1.00 | |
| 0.995 | 7.88 | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 | 2.52 | 2.43 | 2.36 | 2.19 | 2.00 | 1.90 | 1.79 | 1.67 | 1.59 | 1.53 | 1.40 | 1.36 | 1.28 | 1.17 | 1.00 | |

*Compiled from Table A-7c in W. J. Dixon and F. J. Massey, Jr., 1957, *Introduction to Statistical Analysis*, McGraw-Hill Book Company, Inc., New York. Table A-7c has been compiled from values published by Anders Hald, 1952, *Statistical Tables and Formulas*, © 1952, John Wiley & Sons, Inc., New York, and M. Merrington and C. M. Thompson, 1943, "Tables of Percentage Points of the Inverted Beta (F) Distribution," *Biometrika*, 33: 74-87. Reprinted with permission of McGraw-Hill Book Company, Inc., John Wiley & Sons, Inc., the Biometrika Trustees, and Anders Hald.

$$\dagger F_{1-p}(v_1, v_2) = 1/F_p(v_2, v_1).$$

TABLE A7

Pearson and Hartley Charts for the Power of the F-Test*



(Table A7 continued on following pages.)

TABLE A7 (Cont'd)

$\nu_1 = 2$

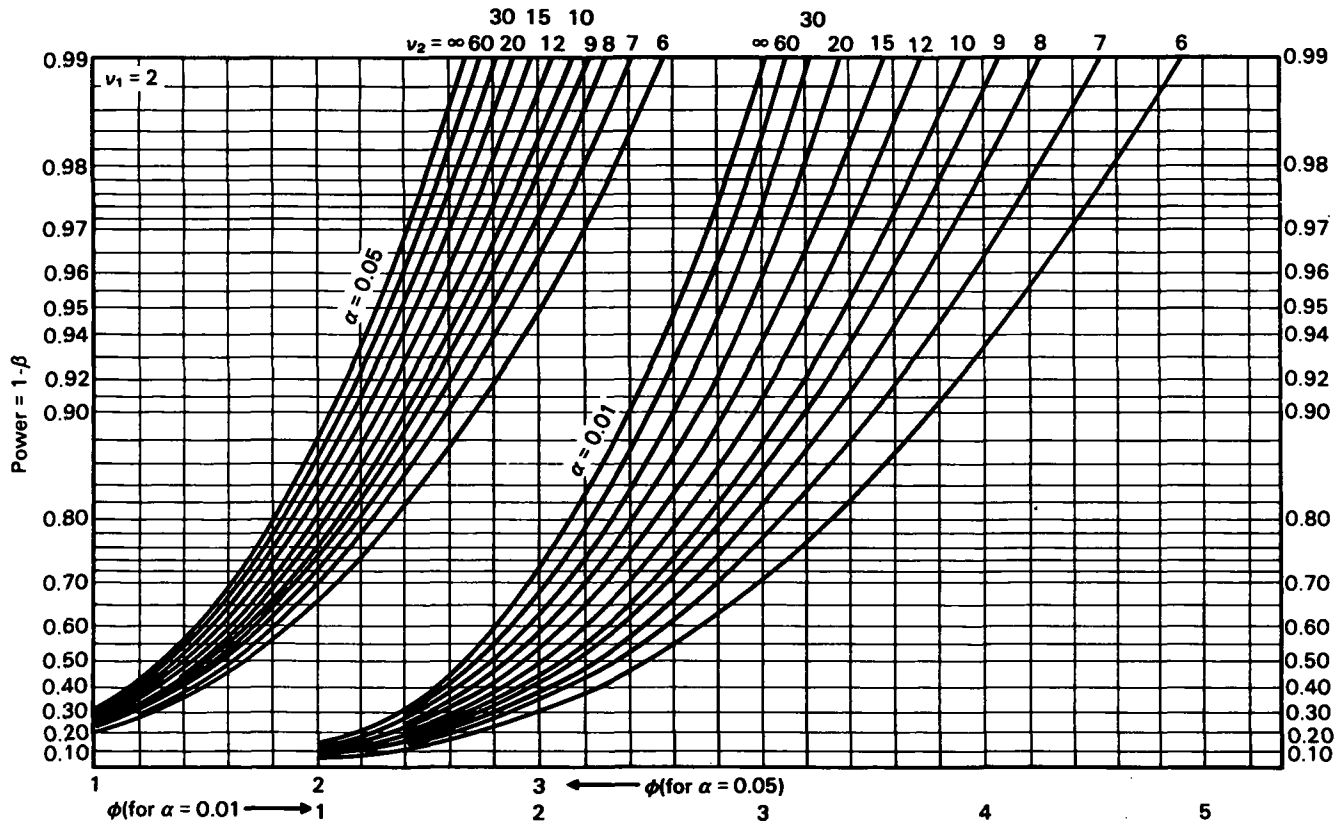
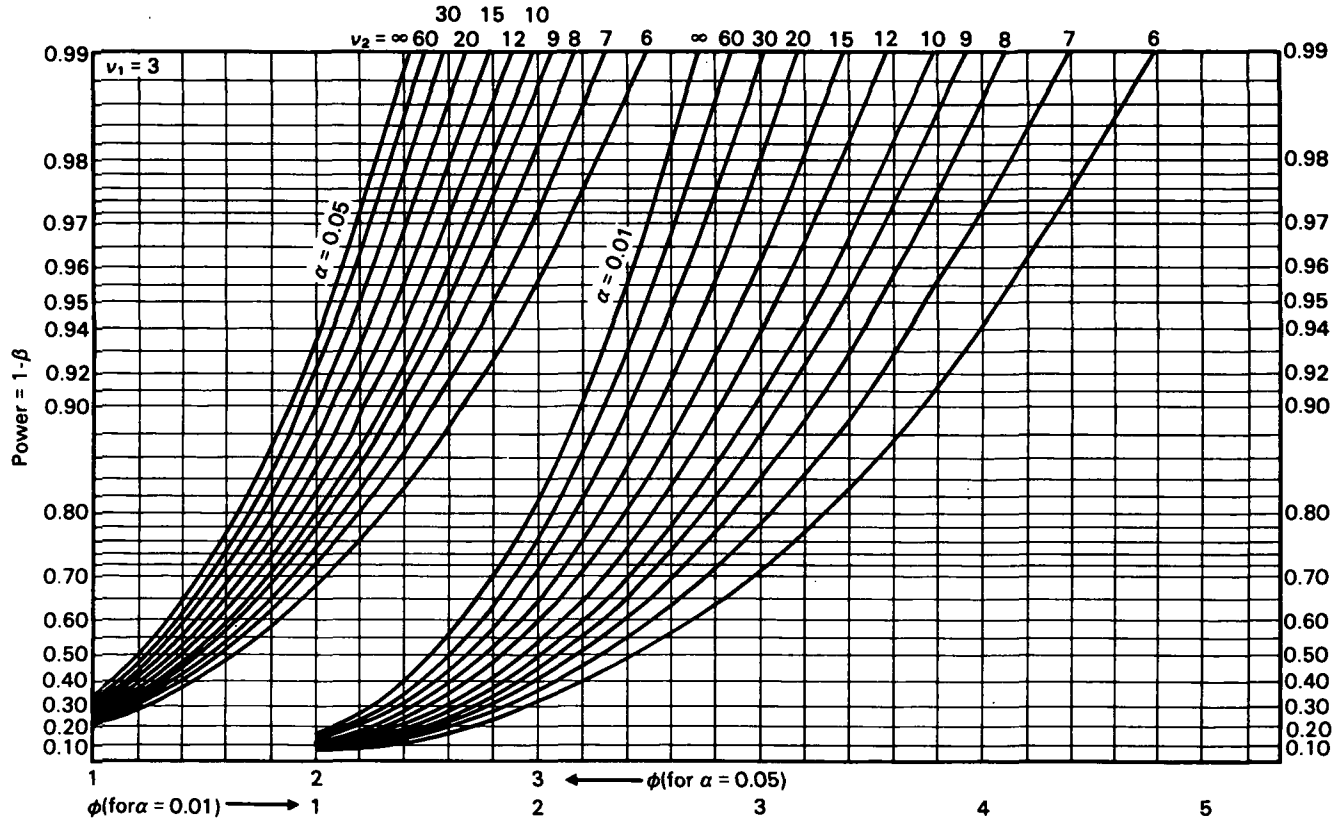


TABLE A7 (Cont'd)

$\nu_1 = 3$



(Table A7 continued on following pages.)

TABLE A7 (Cont'd)

$\nu_1 = 4$

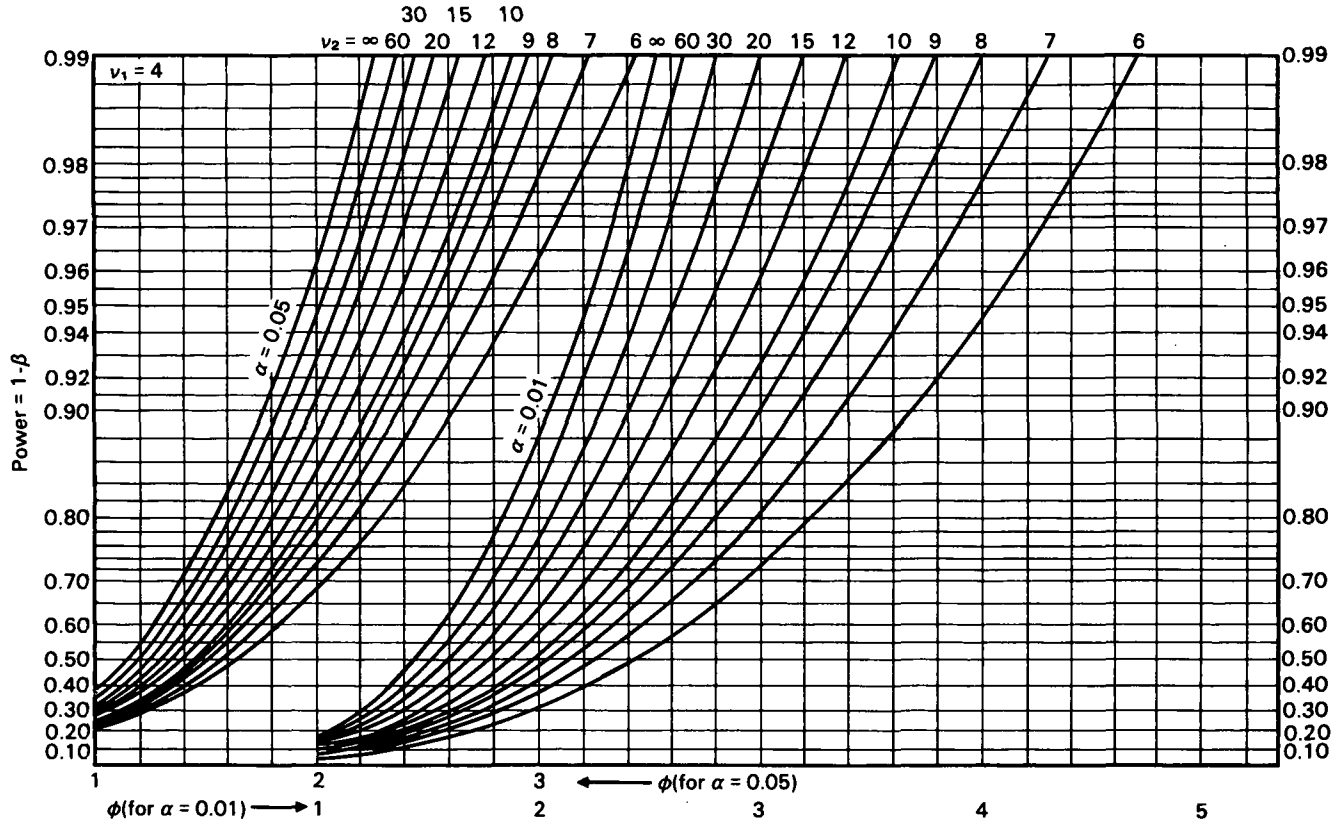
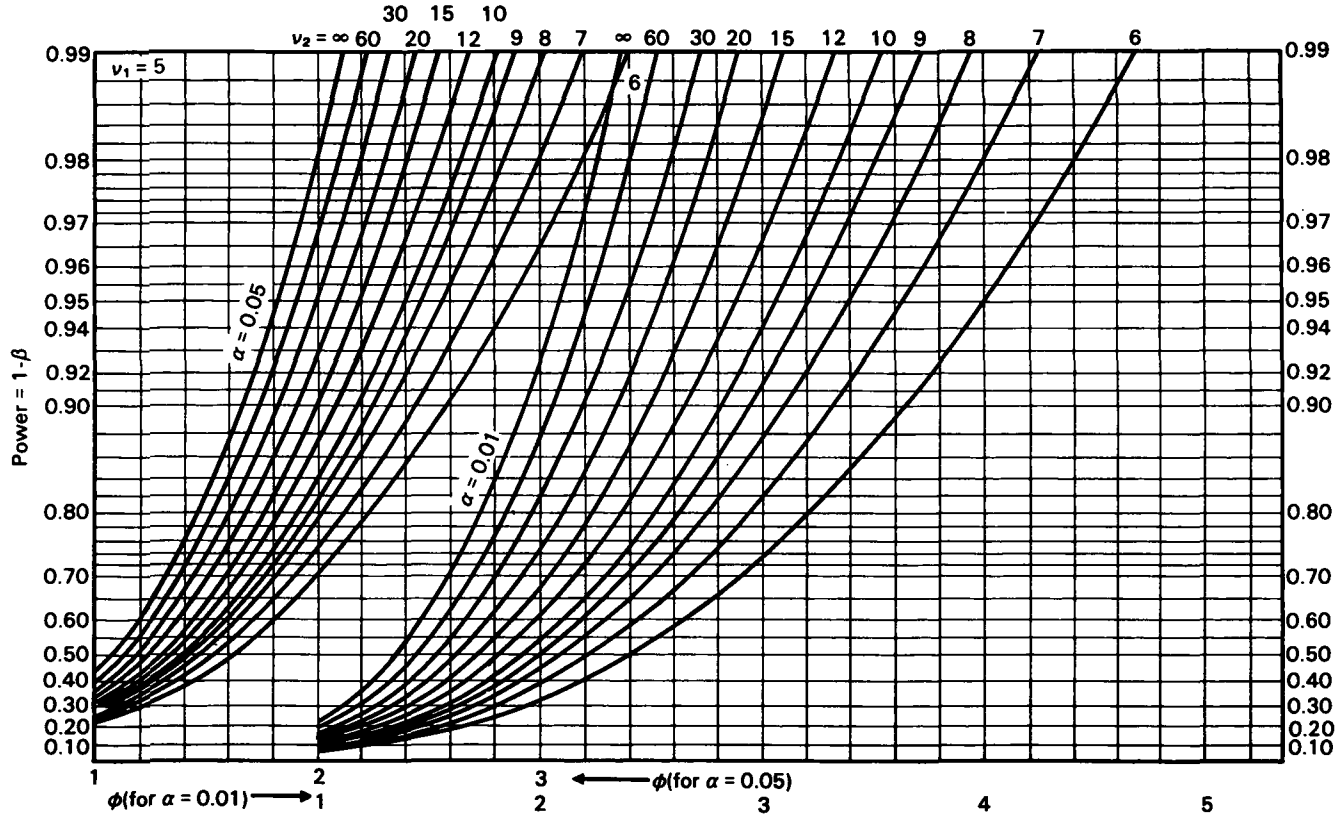


TABLE A7 (Cont'd)

$\nu_1 = 5$



(Table A7 continued on following pages.)

TABLE A7 (Cont'd)

$\nu_1 = 6$

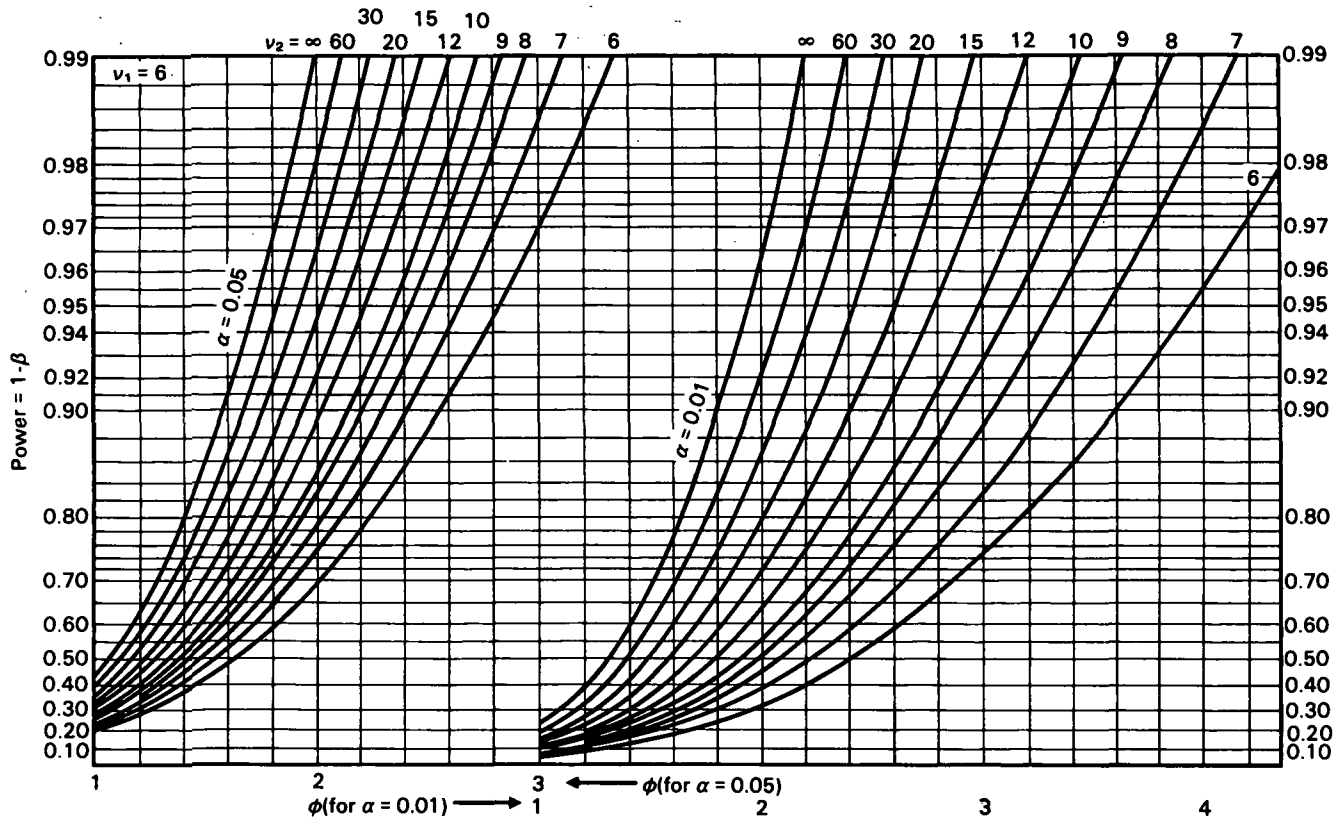
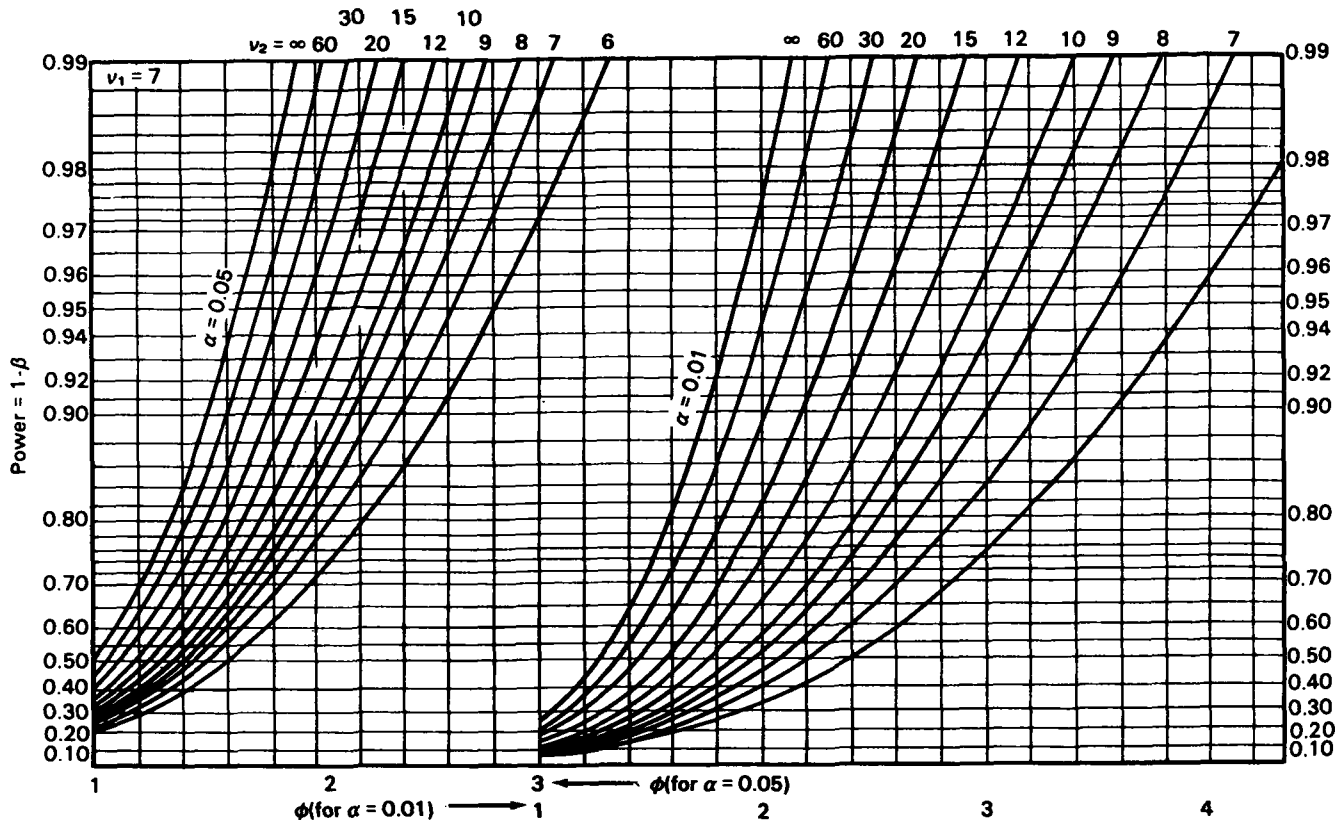


TABLE A7 (Cont'd)

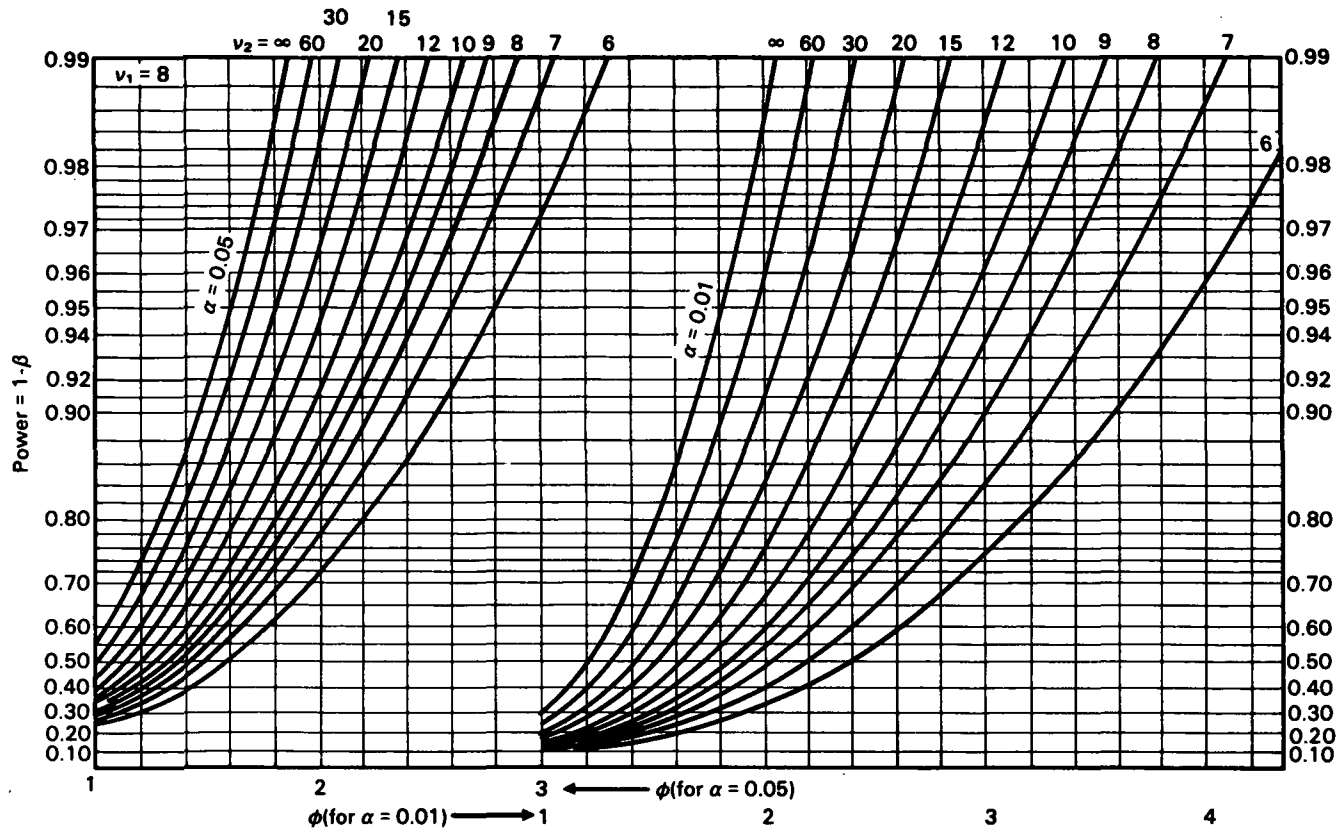
$\nu_1 = 7$



(Table A7 continued on following page.)

TABLE A7 (Cont'd)

$\nu_1 = 8$



*Pearson, E. S., and H. O. Hartley, 1951, *Biometrika*, 38: 115-122. Reproduced with the permission of the Biometrika Trustees.

TABLE A8
Significant Ranges for Duncan's Multiple Range Test at 0.01 Level*

| 0.01 (p, N-a) | | | | | | | | | | | | |
|---------------|------|------|------|------|------|------|------|------|------|------|------|------|
| N-a | p | | | | | | | | | | | |
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 50 | 100 |
| 1 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 |
| 2 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| 3 | 8.26 | 8.5 | 8.6 | 8.7 | 8.8 | 8.9 | 8.9 | 9.0 | 9.0 | 9.3 | 9.3 | 9.3 |
| 4 | 6.51 | 6.8 | 6.9 | 7.0 | 7.1 | 7.1 | 7.2 | 7.2 | 7.3 | 7.5 | 7.5 | 7.5 |
| 5 | 5.70 | 5.96 | 6.11 | 6.18 | 6.26 | 6.33 | 6.40 | 6.44 | 6.5 | 6.8 | 6.8 | 6.8 |
| 6 | 5.24 | 5.51 | 5.65 | 5.73 | 5.81 | 5.88 | 5.95 | 6.00 | 6.0 | 6.3 | 6.3 | 6.3 |
| 7 | 4.95 | 5.22 | 5.37 | 5.45 | 5.53 | 5.61 | 5.69 | 5.73 | 5.8 | 6.0 | 6.0 | 6.0 |
| 8 | 4.74 | 5.00 | 5.14 | 5.23 | 5.32 | 5.40 | 5.47 | 5.51 | 5.5 | 5.8 | 5.8 | 5.8 |
| 9 | 4.60 | 4.86 | 4.99 | 5.08 | 5.17 | 5.25 | 5.32 | 5.36 | 5.4 | 5.7 | 5.7 | 5.7 |
| 10 | 4.48 | 4.73 | 4.88 | 4.96 | 5.06 | 5.13 | 5.20 | 5.24 | 5.28 | 5.55 | 5.55 | 5.55 |
| 11 | 4.39 | 4.63 | 4.77 | 4.86 | 4.94 | 5.01 | 5.06 | 5.12 | 5.15 | 5.39 | 5.39 | 5.39 |
| 12 | 4.32 | 4.55 | 4.68 | 4.76 | 4.84 | 4.92 | 4.96 | 5.02 | 5.07 | 5.26 | 5.26 | 5.26 |
| 13 | 4.26 | 4.48 | 4.62 | 4.69 | 4.74 | 4.84 | 4.88 | 4.94 | 4.98 | 5.15 | 5.15 | 5.15 |
| 14 | 4.21 | 4.42 | 4.55 | 4.63 | 4.70 | 4.78 | 4.83 | 4.87 | 4.91 | 5.07 | 5.07 | 5.07 |
| 15 | 4.17 | 4.37 | 4.50 | 4.58 | 4.64 | 4.72 | 4.77 | 4.81 | 4.84 | 5.00 | 5.00 | 5.00 |
| 16 | 4.13 | 4.34 | 4.45 | 4.54 | 4.60 | 4.67 | 4.72 | 4.76 | 4.79 | 4.94 | 4.94 | 4.94 |
| 17 | 4.10 | 4.30 | 4.41 | 4.50 | 4.56 | 4.63 | 4.68 | 4.73 | 4.75 | 4.89 | 4.89 | 4.89 |
| 18 | 4.07 | 4.27 | 4.38 | 4.46 | 4.53 | 4.59 | 4.64 | 4.68 | 4.71 | 4.85 | 4.85 | 4.85 |
| 19 | 4.05 | 4.24 | 4.35 | 4.43 | 4.50 | 4.56 | 4.61 | 4.64 | 4.67 | 4.82 | 4.82 | 4.82 |
| 20 | 4.02 | 4.22 | 4.33 | 4.40 | 4.47 | 4.53 | 4.58 | 4.61 | 4.65 | 4.79 | 4.79 | 4.79 |
| 30 | 3.89 | 4.06 | 4.16 | 4.22 | 4.32 | 4.36 | 4.41 | 4.45 | 4.48 | 4.65 | 4.71 | 4.71 |
| 40 | 3.82 | 3.99 | 4.10 | 4.17 | 4.24 | 4.30 | 4.34 | 4.37 | 4.41 | 4.59 | 4.69 | 4.69 |
| 60 | 3.76 | 3.92 | 4.03 | 4.12 | 4.17 | 4.23 | 4.27 | 4.31 | 4.34 | 4.53 | 4.66 | 4.66 |
| 100 | 3.71 | 3.86 | 3.98 | 4.06 | 4.11 | 4.17 | 4.21 | 4.25 | 4.29 | 4.48 | 4.64 | 4.65 |
| ∞ | 3.64 | 3.80 | 3.90 | 3.98 | 4.04 | 4.09 | 4.14 | 4.17 | 4.20 | 4.41 | 4.60 | 4.68 |

(Table A8 continued on next page.)

TABLE A8 (Cont'd)
 Significant Ranges for Duncan's Multiple Range Test at 0.05 Level*

| N-a | %0.05 (p, N-a) | | | | | | | | | | | |
|----------|----------------|------|------|------|------|------|------|------|------|------|------|------|
| | p | | | | | | | | | | | |
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 50 | 100 |
| 1 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 |
| 2 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 |
| 3 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 |
| 4 | 3.93 | 4.01 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 |
| 5 | 3.64 | 3.74 | 3.79 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 |
| 6 | 3.46 | 3.58 | 3.64 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 |
| 7 | 3.35 | 3.47 | 3.54 | 3.58 | 3.60 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 |
| 8 | 3.26 | 3.39 | 3.47 | 3.52 | 3.55 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 |
| 9 | 3.20 | 3.34 | 3.41 | 3.47 | 3.50 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 |
| 10 | 3.15 | 3.30 | 3.37 | 3.43 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 |
| 11 | 3.11 | 3.27 | 3.35 | 3.39 | 3.43 | 3.44 | 3.45 | 3.46 | 3.46 | 3.48 | 3.48 | 3.48 |
| 12 | 3.08 | 3.23 | 3.33 | 3.36 | 3.40 | 3.42 | 3.44 | 3.44 | 3.46 | 3.48 | 3.48 | 3.48 |
| 13 | 3.06 | 3.21 | 3.30 | 3.35 | 3.38 | 3.41 | 3.42 | 3.44 | 3.45 | 3.47 | 3.47 | 3.47 |
| 14 | 3.03 | 3.18 | 3.27 | 3.33 | 3.37 | 3.39 | 3.41 | 3.42 | 3.44 | 3.47 | 3.47 | 3.47 |
| 15 | 3.01 | 3.16 | 3.25 | 3.31 | 3.36 | 3.38 | 3.40 | 3.42 | 3.43 | 3.47 | 3.47 | 3.47 |
| 16 | 3.00 | 3.15 | 3.23 | 3.30 | 3.34 | 3.37 | 3.39 | 3.41 | 3.43 | 3.47 | 3.47 | 3.47 |
| 17 | 2.98 | 3.13 | 3.22 | 3.28 | 3.33 | 3.36 | 3.38 | 3.40 | 3.42 | 3.47 | 3.47 | 3.47 |
| 18 | 2.97 | 3.12 | 3.21 | 3.27 | 3.32 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 19 | 2.96 | 3.11 | 3.19 | 3.26 | 3.31 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 20 | 2.95 | 3.10 | 3.18 | 3.25 | 3.30 | 3.34 | 3.36 | 3.38 | 3.40 | 3.47 | 3.47 | 3.47 |
| 30 | 2.89 | 3.04 | 3.12 | 3.20 | 3.25 | 3.29 | 3.32 | 3.35 | 3.37 | 3.47 | 3.47 | 3.47 |
| 40 | 2.86 | 3.01 | 3.10 | 3.17 | 3.22 | 3.27 | 3.30 | 3.33 | 3.35 | 3.47 | 3.47 | 3.47 |
| 60 | 2.83 | 2.98 | 3.08 | 3.14 | 3.20 | 3.24 | 3.28 | 3.31 | 3.33 | 3.47 | 3.48 | 3.48 |
| 100 | 2.80 | 2.95 | 3.05 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.47 | 3.53 | 3.53 |
| ∞ | 2.77 | 2.92 | 3.02 | 3.09 | 3.15 | 3.19 | 3.23 | 3.26 | 3.29 | 3.47 | 3.61 | 3.67 |

*Duncan, D. B., 1955, "Multiple Range and Multiple F Tests," *Biometrics*, 11: 1-42, with the permission of the Biometric Society and the author.

TABLE A9
Upper Bounds for Critical Values for Standardized Residuals*

| n‡ | p† | | | | | | | | | | |
|-----------------|------|------|------|------|------|------|------|------|------|------|--|
| | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 15 | 25 | |
| $\alpha = 0.10$ | | | | | | | | | | | |
| 5 | 1.87 | | | | | | | | | | |
| 6 | 2.00 | 1.89 | | | | | | | | | |
| 7 | 2.10 | 2.02 | 1.90 | | | | | | | | |
| 8 | 2.18 | 2.12 | 2.03 | 1.91 | | | | | | | |
| 9 | 2.24 | 2.20 | 2.13 | 2.05 | 1.92 | | | | | | |
| 10 | 2.30 | 2.26 | 2.21 | 2.15 | 2.06 | 1.92 | | | | | |
| 12 | 2.39 | 2.37 | 2.33 | 2.29 | 2.24 | 2.17 | 1.93 | | | | |
| 14 | 2.47 | 2.45 | 2.42 | 2.39 | 2.36 | 2.32 | 2.19 | 1.94 | | | |
| 16 | 2.53 | 2.51 | 2.50 | 2.47 | 2.45 | 2.42 | 2.34 | 2.20 | | | |
| 18 | 2.58 | 2.57 | 2.56 | 2.54 | 2.52 | 2.50 | 2.44 | 2.25 | | | |
| 20 | 2.63 | 2.62 | 2.61 | 2.59 | 2.58 | 2.56 | 2.52 | 2.46 | 2.11 | | |
| 25 | 2.72 | 2.72 | 2.71 | 2.70 | 2.69 | 2.68 | 2.66 | 2.63 | 2.50 | | |
| 30 | 2.80 | 2.79 | 2.79 | 2.78 | 2.77 | 2.77 | 2.75 | 2.73 | 2.66 | 2.13 | |
| 35 | 2.86 | 2.85 | 2.85 | 2.85 | 2.84 | 2.84 | 2.82 | 2.81 | 2.77 | 2.55 | |
| 40 | 2.91 | 2.91 | 2.90 | 2.90 | 2.90 | 2.89 | 2.88 | 2.87 | 2.84 | 2.72 | |
| 45 | 2.95 | 2.95 | 2.95 | 2.95 | 2.94 | 2.94 | 2.93 | 2.93 | 2.90 | 2.82 | |
| 50 | 2.99 | 2.99 | 2.99 | 2.99 | 2.98 | 2.98 | 2.98 | 2.97 | 2.95 | 2.89 | |
| 60 | 3.06 | 3.06 | 3.05 | 3.05 | 3.05 | 3.05 | 3.05 | 3.04 | 3.03 | 3.00 | |
| 70 | 3.11 | 3.11 | 3.11 | 3.11 | 3.11 | 3.11 | 3.10 | 3.10 | 3.09 | 3.07 | |
| 80 | 3.16 | 3.16 | 3.16 | 3.15 | 3.15 | 3.15 | 3.15 | 3.15 | 3.14 | 3.12 | |
| 90 | 3.20 | 3.20 | 3.19 | 3.19 | 3.19 | 3.19 | 3.19 | 3.19 | 3.18 | 3.17 | |
| 100 | 3.23 | 3.23 | 3.23 | 3.23 | 3.23 | 3.23 | 3.23 | 3.22 | 3.22 | 3.21 | |
| $\alpha = 0.05$ | | | | | | | | | | | |
| 5 | 1.92 | | | | | | | | | | |
| 6 | 2.07 | 1.93 | | | | | | | | | |
| 7 | 2.19 | 2.08 | 1.94 | | | | | | | | |
| 8 | 2.28 | 2.20 | 2.10 | 1.94 | | | | | | | |
| 9 | 2.35 | 2.29 | 2.21 | 2.10 | 1.95 | | | | | | |
| 10 | 2.42 | 2.37 | 2.31 | 2.22 | 2.11 | 1.95 | | | | | |
| 12 | 2.52 | 2.49 | 2.45 | 2.39 | 2.33 | 2.24 | 1.96 | | | | |
| 14 | 2.61 | 2.58 | 2.55 | 2.51 | 2.47 | 2.41 | 2.25 | 1.96 | | | |
| 16 | 2.68 | 2.66 | 2.63 | 2.60 | 2.57 | 2.53 | 2.43 | 2.26 | | | |
| 18 | 2.73 | 2.72 | 2.70 | 2.68 | 2.65 | 2.62 | 2.55 | 2.44 | | | |
| 20 | 2.78 | 2.77 | 2.76 | 2.74 | 2.72 | 2.70 | 2.64 | 2.57 | 2.15 | | |
| 25 | 2.89 | 2.88 | 2.87 | 2.86 | 2.84 | 2.83 | 2.80 | 2.76 | 2.60 | | |
| 30 | 2.96 | 2.96 | 2.95 | 2.94 | 2.93 | 2.93 | 2.90 | 2.88 | 2.79 | 2.17 | |
| 35 | 3.03 | 3.02 | 3.02 | 3.01 | 3.00 | 3.00 | 2.93 | 2.97 | 2.91 | 2.64 | |
| 40 | 3.08 | 3.08 | 3.07 | 3.07 | 3.06 | 3.06 | 3.05 | 3.03 | 3.00 | 2.84 | |
| 45 | 3.13 | 3.12 | 3.12 | 3.12 | 3.11 | 3.11 | 3.10 | 3.09 | 3.06 | 2.96 | |
| 50 | 3.17 | 3.16 | 3.16 | 3.16 | 3.15 | 3.15 | 3.14 | 3.14 | 3.11 | 3.04 | |
| 60 | 3.23 | 3.23 | 3.23 | 3.23 | 3.22 | 3.22 | 3.22 | 3.21 | 3.20 | 3.15 | |
| 70 | 3.29 | 3.29 | 3.28 | 3.28 | 3.28 | 3.28 | 3.27 | 3.27 | 3.26 | 3.23 | |
| 80 | 3.33 | 3.33 | 3.33 | 3.33 | 3.33 | 3.33 | 3.32 | 3.32 | 3.31 | 3.29 | |
| 90 | 3.37 | 3.37 | 3.37 | 3.37 | 3.37 | 3.37 | 3.36 | 3.36 | 3.36 | 3.34 | |
| 100 | 3.41 | 3.41 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.40 | 3.39 | 3.38 | |

(Table A9 continued on next page.)

TABLE A9 (Cont'd)

| p^\dagger | | | | | | | | | | |
|-----------------|------|------|------|------|------|------|------|------|------|------|
| n^\ddagger | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 15 | 25 |
| $\alpha = 0.01$ | | | | | | | | | | |
| 5 | 1.98 | | | | | | | | | |
| 6 | 2.17 | 1.98 | | | | | | | | |
| 7 | 2.32 | 2.17 | 1.98 | | | | | | | |
| 8 | 2.44 | 2.32 | 2.18 | 1.98 | | | | | | |
| 9 | 2.54 | 2.44 | 2.33 | 2.18 | 1.99 | | | | | |
| 10 | 2.62 | 2.55 | 2.45 | 2.33 | 2.18 | 1.99 | | | | |
| 12 | 2.76 | 2.70 | 2.64 | 2.56 | 2.46 | 2.34 | 1.99 | | | |
| 14 | 2.86 | 2.82 | 2.78 | 2.72 | 2.65 | 2.57 | 2.35 | 1.99 | | |
| 16 | 2.95 | 2.92 | 2.88 | 2.84 | 2.79 | 2.73 | 2.58 | 2.35 | | |
| 18 | 3.02 | 3.00 | 2.97 | 2.94 | 2.90 | 2.85 | 2.75 | 2.59 | | |
| 20 | 3.08 | 3.06 | 3.04 | 3.01 | 2.98 | 2.95 | 2.87 | 2.76 | 2.20 | |
| 25 | 3.21 | 3.19 | 3.18 | 3.16 | 3.14 | 3.12 | 3.07 | 3.01 | 2.78 | |
| 30 | 3.30 | 3.29 | 3.28 | 3.26 | 3.25 | 3.24 | 3.21 | 3.17 | 3.04 | 2.21 |
| 35 | 3.37 | 3.36 | 3.35 | 3.34 | 3.34 | 3.33 | 3.30 | 3.28 | 3.19 | 2.81 |
| 40 | 3.43 | 3.42 | 3.42 | 3.41 | 3.40 | 3.40 | 3.38 | 3.36 | 3.30 | 3.08 |
| 45 | 3.48 | 3.47 | 3.47 | 3.46 | 3.46 | 3.45 | 3.44 | 3.43 | 3.38 | 3.23 |
| 50 | 3.52 | 3.52 | 3.51 | 3.51 | 3.51 | 3.50 | 3.49 | 3.48 | 3.45 | 3.34 |
| 60 | 3.60 | 3.59 | 3.59 | 3.59 | 3.58 | 3.58 | 3.57 | 3.56 | 3.54 | 3.48 |
| 70 | 3.65 | 3.65 | 3.65 | 3.65 | 3.64 | 3.64 | 3.64 | 3.63 | 3.61 | 3.57 |
| 80 | 3.70 | 3.70 | 3.70 | 3.70 | 3.69 | 3.69 | 3.69 | 3.68 | 3.67 | 3.64 |
| 90 | 3.74 | 3.74 | 3.74 | 3.74 | 3.74 | 3.74 | 3.73 | 3.73 | 3.72 | 3.70 |
| 100 | 3.78 | 3.78 | 3.78 | 3.77 | 3.77 | 3.77 | 3.77 | 3.77 | 3.76 | 3.74 |

*Reprinted from Table 1 in Richard E. Lund, 1975, "Tables for Approximate Test for Outliers in Linear Models," *Technometrics*, 17: 473-476, with the permission of the publisher and the author.

$^\dagger p$ = number of independent variables (including intercept if fitted).

$^\ddagger n$ = number of observations.

TABLE A10
Durbin-Watson Critical Values*

| n | k = 1 | | k = 2 | | k = 3 | | k = 4 | | k = 5 | |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U |
| $\alpha = 0.05$ | | | | | | | | | | |
| 15 | 1.08 | 1.36 | 0.95 | 1.54 | 0.82 | 1.75 | 0.69 | 1.97 | 0.56 | 2.21 |
| 16 | 1.10 | 1.37 | 0.98 | 1.54 | 0.86 | 1.73 | 0.74 | 1.93 | 0.62 | 2.15 |
| 17 | 1.13 | 1.38 | 1.02 | 1.54 | 0.90 | 1.71 | 0.78 | 1.90 | 0.67 | 2.10 |
| 18 | 1.16 | 1.39 | 1.05 | 1.53 | 0.93 | 1.69 | 0.82 | 1.87 | 0.71 | 2.06 |
| 19 | 1.18 | 1.40 | 1.08 | 1.53 | 0.97 | 1.68 | 0.86 | 1.85 | 0.75 | 2.02 |
| 20 | 1.20 | 1.41 | 1.10 | 1.54 | 1.00 | 1.68 | 0.90 | 1.83 | 0.79 | 1.99 |
| 21 | 1.22 | 1.42 | 1.13 | 1.54 | 1.03 | 1.67 | 0.93 | 1.81 | 0.83 | 1.96 |
| 22 | 1.24 | 1.43 | 1.15 | 1.54 | 1.05 | 1.66 | 0.96 | 1.80 | 0.86 | 1.94 |
| 23 | 1.26 | 1.44 | 1.17 | 1.54 | 1.08 | 1.66 | 0.99 | 1.79 | 0.90 | 1.92 |
| 24 | 1.27 | 1.45 | 1.19 | 1.55 | 1.10 | 1.66 | 1.01 | 1.78 | 0.93 | 1.90 |
| 25 | 1.29 | 1.45 | 1.21 | 1.55 | 1.12 | 1.66 | 1.04 | 1.77 | 0.95 | 1.89 |
| 26 | 1.30 | 1.46 | 1.22 | 1.55 | 1.14 | 1.65 | 1.06 | 1.76 | 0.98 | 1.88 |
| 27 | 1.32 | 1.47 | 1.24 | 1.56 | 1.16 | 1.65 | 1.08 | 1.76 | 1.01 | 1.86 |
| 28 | 1.33 | 1.48 | 1.26 | 1.56 | 1.18 | 1.65 | 1.10 | 1.75 | 1.03 | 1.85 |
| 29 | 1.34 | 1.48 | 1.27 | 1.56 | 1.20 | 1.65 | 1.12 | 1.74 | 1.05 | 1.84 |
| 30 | 1.35 | 1.49 | 1.28 | 1.57 | 1.21 | 1.65 | 1.14 | 1.74 | 1.07 | 1.83 |
| 31 | 1.36 | 1.50 | 1.30 | 1.57 | 1.23 | 1.65 | 1.16 | 1.74 | 1.09 | 1.83 |
| 32 | 1.37 | 1.50 | 1.31 | 1.57 | 1.24 | 1.65 | 1.18 | 1.73 | 1.11 | 1.82 |
| 33 | 1.38 | 1.51 | 1.32 | 1.58 | 1.26 | 1.65 | 1.19 | 1.73 | 1.13 | 1.81 |
| 34 | 1.39 | 1.51 | 1.33 | 1.58 | 1.27 | 1.65 | 1.21 | 1.73 | 1.15 | 1.81 |
| 35 | 1.40 | 1.52 | 1.34 | 1.58 | 1.28 | 1.65 | 1.22 | 1.73 | 1.16 | 1.80 |
| 36 | 1.41 | 1.52 | 1.35 | 1.59 | 1.29 | 1.65 | 1.24 | 1.73 | 1.18 | 1.80 |
| 37 | 1.42 | 1.53 | 1.36 | 1.59 | 1.31 | 1.66 | 1.25 | 1.72 | 1.19 | 1.80 |
| 38 | 1.43 | 1.54 | 1.37 | 1.59 | 1.32 | 1.66 | 1.26 | 1.72 | 1.21 | 1.79 |
| 39 | 1.43 | 1.54 | 1.38 | 1.60 | 1.33 | 1.66 | 1.27 | 1.72 | 1.22 | 1.79 |
| 40 | 1.44 | 1.54 | 1.39 | 1.60 | 1.34 | 1.66 | 1.29 | 1.72 | 1.23 | 1.79 |
| 45 | 1.48 | 1.57 | 1.43 | 1.62 | 1.38 | 1.67 | 1.34 | 1.72 | 1.29 | 1.78 |
| 50 | 1.50 | 1.59 | 1.46 | 1.63 | 1.42 | 1.67 | 1.38 | 1.72 | 1.34 | 1.77 |
| 55 | 1.53 | 1.60 | 1.49 | 1.64 | 1.45 | 1.68 | 1.41 | 1.72 | 1.38 | 1.77 |
| 60 | 1.55 | 1.62 | 1.51 | 1.65 | 1.48 | 1.69 | 1.44 | 1.73 | 1.41 | 1.77 |
| 65 | 1.57 | 1.63 | 1.54 | 1.66 | 1.50 | 1.70 | 1.47 | 1.73 | 1.44 | 1.77 |
| 70 | 1.58 | 1.64 | 1.55 | 1.67 | 1.52 | 1.70 | 1.49 | 1.74 | 1.46 | 1.77 |
| 75 | 1.60 | 1.65 | 1.57 | 1.68 | 1.54 | 1.71 | 1.51 | 1.74 | 1.49 | 1.77 |
| 80 | 1.61 | 1.66 | 1.59 | 1.69 | 1.56 | 1.72 | 1.53 | 1.74 | 1.51 | 1.77 |
| 85 | 1.62 | 1.67 | 1.60 | 1.70 | 1.57 | 1.72 | 1.55 | 1.75 | 1.52 | 1.77 |
| 90 | 1.63 | 1.68 | 1.61 | 1.70 | 1.59 | 1.73 | 1.57 | 1.75 | 1.54 | 1.78 |
| 95 | 1.64 | 1.69 | 1.62 | 1.71 | 1.60 | 1.73 | 1.58 | 1.75 | 1.56 | 1.78 |
| 100 | 1.65 | 1.69 | 1.63 | 1.72 | 1.61 | 1.74 | 1.59 | 1.76 | 1.57 | 1.78 |

(Table A10 continued on next page).

TABLE A10 (Cont'd)

| n | k = 1 | | k = 2 | | k = 3 | | k = 4 | | k = 5 | |
|------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U |
| $\alpha = 0.025$ | | | | | | | | | | |
| 15 | 0.95 | 1.23 | 0.83 | 1.40 | 0.71 | 1.61 | 0.59 | 1.84 | 0.48 | 2.09 |
| 16 | 0.98 | 1.24 | 0.86 | 1.40 | 0.75 | 1.59 | 0.64 | 1.80 | 0.53 | 2.03 |
| 17 | 1.01 | 1.25 | 0.90 | 1.40 | 0.79 | 1.58 | 0.68 | 1.77 | 0.57 | 1.98 |
| 18 | 1.03 | 1.26 | 0.93 | 1.40 | 0.82 | 1.56 | 0.72 | 1.74 | 0.62 | 1.93 |
| 19 | 1.06 | 1.28 | 0.96 | 1.41 | 0.86 | 1.55 | 0.76 | 1.72 | 0.66 | 1.90 |
| 20 | 1.08 | 1.28 | 0.99 | 1.41 | 0.89 | 1.55 | 0.79 | 1.70 | 0.70 | 1.87 |
| 21 | 1.10 | 1.30 | 1.01 | 1.41 | 0.92 | 1.54 | 0.83 | 1.69 | 0.73 | 1.84 |
| 22 | 1.12 | 1.31 | 1.04 | 1.42 | 0.95 | 1.54 | 0.86 | 1.68 | 0.77 | 1.82 |
| 23 | 1.14 | 1.32 | 1.06 | 1.42 | 0.97 | 1.54 | 0.89 | 1.67 | 0.80 | 1.80 |
| 24 | 1.16 | 1.33 | 1.08 | 1.43 | 1.00 | 1.54 | 0.91 | 1.66 | 0.83 | 1.79 |
| 25 | 1.18 | 1.34 | 1.10 | 1.43 | 1.02 | 1.54 | 0.94 | 1.65 | 0.86 | 1.77 |
| 26 | 1.19 | 1.35 | 1.12 | 1.44 | 1.04 | 1.54 | 0.96 | 1.65 | 0.88 | 1.76 |
| 27 | 1.21 | 1.36 | 1.13 | 1.44 | 1.06 | 1.54 | 0.99 | 1.64 | 0.91 | 1.75 |
| 28 | 1.22 | 1.37 | 1.15 | 1.45 | 1.08 | 1.54 | 1.01 | 1.64 | 0.93 | 1.74 |
| 29 | 1.24 | 1.38 | 1.17 | 1.45 | 1.10 | 1.54 | 1.03 | 1.63 | 0.96 | 1.73 |
| 30 | 1.25 | 1.38 | 1.18 | 1.46 | 1.12 | 1.54 | 1.05 | 1.63 | 0.98 | 1.73 |
| 31 | 1.26 | 1.39 | 1.20 | 1.47 | 1.13 | 1.55 | 1.07 | 1.63 | 1.00 | 1.72 |
| 32 | 1.27 | 1.40 | 1.21 | 1.47 | 1.15 | 1.55 | 1.08 | 1.63 | 1.02 | 1.71 |
| 33 | 1.28 | 1.41 | 1.22 | 1.48 | 1.16 | 1.55 | 1.10 | 1.63 | 1.04 | 1.71 |
| 34 | 1.29 | 1.41 | 1.24 | 1.48 | 1.17 | 1.55 | 1.12 | 1.63 | 1.06 | 1.70 |
| 35 | 1.30 | 1.42 | 1.25 | 1.48 | 1.19 | 1.55 | 1.13 | 1.63 | 1.07 | 1.70 |
| 36 | 1.31 | 1.43 | 1.26 | 1.49 | 1.20 | 1.56 | 1.15 | 1.63 | 1.09 | 1.70 |
| 37 | 1.32 | 1.43 | 1.27 | 1.49 | 1.21 | 1.56 | 1.16 | 1.62 | 1.10 | 1.70 |
| 38 | 1.33 | 1.44 | 1.28 | 1.50 | 1.23 | 1.56 | 1.17 | 1.62 | 1.12 | 1.70 |
| 39 | 1.34 | 1.44 | 1.29 | 1.50 | 1.24 | 1.56 | 1.19 | 1.63 | 1.13 | 1.69 |
| 40 | 1.35 | 1.45 | 1.30 | 1.51 | 1.25 | 1.57 | 1.20 | 1.63 | 1.15 | 1.69 |
| 45 | 1.39 | 1.48 | 1.34 | 1.53 | 1.30 | 1.58 | 1.25 | 1.63 | 1.21 | 1.69 |
| 50 | 1.42 | 1.50 | 1.38 | 1.54 | 1.34 | 1.59 | 1.30 | 1.64 | 1.26 | 1.69 |
| 55 | 1.45 | 1.52 | 1.41 | 1.56 | 1.37 | 1.60 | 1.33 | 1.64 | 1.30 | 1.69 |
| 60 | 1.47 | 1.54 | 1.44 | 1.57 | 1.40 | 1.61 | 1.37 | 1.65 | 1.33 | 1.69 |
| 65 | 1.49 | 1.55 | 1.46 | 1.59 | 1.43 | 1.62 | 1.40 | 1.66 | 1.36 | 1.69 |
| 70 | 1.51 | 1.57 | 1.48 | 1.60 | 1.45 | 1.63 | 1.42 | 1.66 | 1.39 | 1.70 |
| 75 | 1.53 | 1.58 | 1.50 | 1.61 | 1.47 | 1.64 | 1.45 | 1.67 | 1.42 | 1.70 |
| 80 | 1.54 | 1.59 | 1.52 | 1.69 | 1.49 | 1.65 | 1.47 | 1.67 | 1.44 | 1.70 |
| 85 | 1.56 | 1.60 | 1.53 | 1.63 | 1.51 | 1.65 | 1.49 | 1.68 | 1.46 | 1.71 |
| 90 | 1.57 | 1.61 | 1.55 | 1.64 | 1.53 | 1.66 | 1.50 | 1.69 | 1.48 | 1.71 |
| 95 | 1.58 | 1.62 | 1.56 | 1.65 | 1.54 | 1.67 | 1.52 | 1.69 | 1.50 | 1.71 |
| 100 | 1.59 | 1.63 | 1.57 | 1.65 | 1.55 | 1.67 | 1.53 | 1.70 | 1.51 | 1.72 |

(Table A10 continued on next page).

TABLE A10 (Cont'd)

| n | k = 1 | | k = 2 | | k = 3 | | k = 4 | | k = 5 | |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U | d _L | d _U |
| $\alpha = 0.01$ | | | | | | | | | | |
| 15 | 0.81 | 1.07 | 0.70 | 1.25 | 0.59 | 1.46 | 0.49 | 1.70 | 0.39 | 1.96 |
| 16 | 0.84 | 1.09 | 0.74 | 1.25 | 0.63 | 1.44 | 0.53 | 1.66 | 0.44 | 1.90 |
| 17 | 0.87 | 1.10 | 0.77 | 1.25 | 0.67 | 1.43 | 0.57 | 1.63 | 0.48 | 1.85 |
| 18 | 0.90 | 1.12 | 0.80 | 1.26 | 0.71 | 1.42 | 0.61 | 1.60 | 0.52 | 1.80 |
| 19 | 0.93 | 1.13 | 0.83 | 1.26 | 0.74 | 1.41 | 0.65 | 1.58 | 0.56 | 1.77 |
| 20 | 0.95 | 1.15 | 0.86 | 1.27 | 0.77 | 1.41 | 0.68 | 1.57 | 0.60 | 1.74 |
| 21 | 0.97 | 1.16 | 0.89 | 1.27 | 0.80 | 1.41 | 0.72 | 1.55 | 0.63 | 1.71 |
| 22 | 1.00 | 1.17 | 0.91 | 1.28 | 0.83 | 1.40 | 0.75 | 1.54 | 0.66 | 1.69 |
| 23 | 1.02 | 1.19 | 0.94 | 1.29 | 0.86 | 1.40 | 0.77 | 1.53 | 0.70 | 1.67 |
| 24 | 1.04 | 1.20 | 0.96 | 1.30 | 0.88 | 1.41 | 0.80 | 1.53 | 0.72 | 1.66 |
| 25 | 1.05 | 1.21 | 0.98 | 1.30 | 0.90 | 1.41 | 0.83 | 1.52 | 0.75 | 1.65 |
| 26 | 1.07 | 1.22 | 1.00 | 1.31 | 0.93 | 1.41 | 0.85 | 1.52 | 0.78 | 1.64 |
| 27 | 1.09 | 1.23 | 1.02 | 1.32 | 0.95 | 1.41 | 0.88 | 1.51 | 0.81 | 1.63 |
| 28 | 1.10 | 1.24 | 1.04 | 1.32 | 0.97 | 1.41 | 0.90 | 1.51 | 0.83 | 1.62 |
| 29 | 1.12 | 1.25 | 1.05 | 1.33 | 0.99 | 1.42 | 0.92 | 1.51 | 0.85 | 1.61 |
| 30 | 1.13 | 1.26 | 1.07 | 1.34 | 1.01 | 1.42 | 0.94 | 1.51 | 0.88 | 1.61 |
| 31 | 1.15 | 1.27 | 1.08 | 1.34 | 1.02 | 1.42 | 0.96 | 1.51 | 0.90 | 1.60 |
| 32 | 1.16 | 1.28 | 1.10 | 1.35 | 1.04 | 1.43 | 0.98 | 1.51 | 0.92 | 1.60 |
| 33 | 1.17 | 1.29 | 1.11 | 1.36 | 1.05 | 1.43 | 1.00 | 1.51 | 0.94 | 1.59 |
| 34 | 1.18 | 1.30 | 1.13 | 1.36 | 1.07 | 1.43 | 1.01 | 1.51 | 0.95 | 1.59 |
| 35 | 1.19 | 1.31 | 1.14 | 1.37 | 1.08 | 1.44 | 1.03 | 1.51 | 0.97 | 1.59 |
| 36 | 1.21 | 1.32 | 1.15 | 1.38 | 1.10 | 1.44 | 1.04 | 1.51 | 0.99 | 1.59 |
| 37 | 1.22 | 1.32 | 1.16 | 1.38 | 1.11 | 1.45 | 1.06 | 1.51 | 1.00 | 1.59 |
| 38 | 1.23 | 1.33 | 1.18 | 1.39 | 1.12 | 1.45 | 1.07 | 1.52 | 1.02 | 1.58 |
| 39 | 1.24 | 1.34 | 1.19 | 1.39 | 1.14 | 1.45 | 1.09 | 1.52 | 1.03 | 1.58 |
| 40 | 1.25 | 1.34 | 1.20 | 1.40 | 1.15 | 1.46 | 1.10 | 1.52 | 1.05 | 1.58 |
| 45 | 1.29 | 1.38 | 1.24 | 1.42 | 1.20 | 1.48 | 1.16 | 1.53 | 1.11 | 1.58 |
| 50 | 1.32 | 1.40 | 1.28 | 1.45 | 1.24 | 1.49 | 1.20 | 1.54 | 1.16 | 1.59 |
| 55 | 1.36 | 1.43 | 1.32 | 1.47 | 1.28 | 1.51 | 1.25 | 1.55 | 1.21 | 1.59 |
| 60 | 1.38 | 1.45 | 1.35 | 1.48 | 1.32 | 1.52 | 1.28 | 1.56 | 1.25 | 1.60 |
| 65 | 1.41 | 1.47 | 1.38 | 1.50 | 1.35 | 1.53 | 1.31 | 1.57 | 1.28 | 1.61 |
| 70 | 1.43 | 1.49 | 1.40 | 1.52 | 1.37 | 1.55 | 1.34 | 1.58 | 1.31 | 1.61 |
| 75 | 1.45 | 1.50 | 1.42 | 1.53 | 1.39 | 1.56 | 1.37 | 1.59 | 1.34 | 1.62 |
| 80 | 1.47 | 1.52 | 1.44 | 1.54 | 1.42 | 1.57 | 1.39 | 1.60 | 1.36 | 1.62 |
| 85 | 1.48 | 1.53 | 1.46 | 1.55 | 1.43 | 1.58 | 1.41 | 1.60 | 1.39 | 1.63 |
| 90 | 1.50 | 1.54 | 1.47 | 1.56 | 1.45 | 1.59 | 1.43 | 1.61 | 1.41 | 1.64 |
| 95 | 1.51 | 1.55 | 1.49 | 1.57 | 1.47 | 1.60 | 1.45 | 1.62 | 1.42 | 1.64 |
| 100 | 1.52 | 1.56 | 1.50 | 1.58 | 1.48 | 1.60 | 1.46 | 1.63 | 1.44 | 1.65 |

*Durbin, J., and G. S. Watson, 1951, *Biometrika*, 38: 159-178. Reprinted with the permission of the Biometrika Trustees and the principal author.

TABLE A11

Cumulative Probabilities $\Pr\{U \leq u\}$ of the Total Number of Runs u in a Sample of Size $n_1 + n_2$,
Where n_1 Is the Number of Plus Signs and n_2 Is the Number of Minus Signs*†

| $(n_1, n_2) U$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (2, 3) | 0.200 | 0.500 | 0.900 | 1.000 | | | | | | | | | | | | | | | |
| (2, 4) | 0.133 | 0.400 | 0.800 | 1.000 | | | | | | | | | | | | | | | |
| (2, 5) | 0.095 | 0.333 | 0.714 | 1.000 | | | | | | | | | | | | | | | |
| (2, 6) | 0.071 | 0.286 | 0.643 | 1.000 | | | | | | | | | | | | | | | |
| (2, 7) | 0.056 | 0.250 | 0.583 | 1.000 | | | | | | | | | | | | | | | |
| (2, 8) | 0.044 | 0.222 | 0.533 | 1.000 | | | | | | | | | | | | | | | |
| (2, 9) | 0.036 | 0.200 | 0.491 | 1.000 | | | | | | | | | | | | | | | |
| (2, 10) | 0.030 | 0.182 | 0.455 | 1.000 | | | | | | | | | | | | | | | |
| (3, 3) | 0.100 | 0.300 | 0.700 | 0.900 | 1.000 | | | | | | | | | | | | | | |
| (3, 4) | 0.057 | 0.200 | 0.543 | 0.800 | 0.971 | 1.000 | | | | | | | | | | | | | |
| (3, 5) | 0.036 | 0.143 | 0.429 | 0.714 | 0.929 | 1.000 | | | | | | | | | | | | | |
| (3, 6) | 0.024 | 0.107 | 0.345 | 0.643 | 0.881 | 1.000 | | | | | | | | | | | | | |
| (3, 7) | 0.017 | 0.083 | 0.283 | 0.583 | 0.833 | 1.000 | | | | | | | | | | | | | |
| (3, 8) | 0.012 | 0.067 | 0.236 | 0.533 | 0.788 | 1.000 | | | | | | | | | | | | | |
| (3, 9) | 0.009 | 0.055 | 0.200 | 0.491 | 0.745 | 1.000 | | | | | | | | | | | | | |
| (3, 10) | 0.007 | 0.045 | 0.171 | 0.455 | 0.706 | 1.000 | | | | | | | | | | | | | |
| (4, 4) | 0.029 | 0.114 | 0.371 | 0.629 | 0.886 | 0.971 | 1.000 | | | | | | | | | | | | |
| (4, 5) | 0.016 | 0.071 | 0.262 | 0.500 | 0.786 | 0.929 | 0.992 | 1.000 | | | | | | | | | | | |
| (4, 6) | 0.010 | 0.048 | 0.190 | 0.405 | 0.690 | 0.881 | 0.976 | 1.000 | | | | | | | | | | | |
| (4, 7) | 0.006 | 0.033 | 0.142 | 0.333 | 0.606 | 0.833 | 0.954 | 1.000 | | | | | | | | | | | |
| (4, 8) | 0.004 | 0.024 | 0.109 | 0.279 | 0.533 | 0.788 | 0.929 | 1.000 | | | | | | | | | | | |
| (4, 9) | 0.003 | 0.018 | 0.085 | 0.236 | 0.471 | 0.745 | 0.902 | 1.000 | | | | | | | | | | | |
| (4, 10) | 0.002 | 0.014 | 0.068 | 0.203 | 0.419 | 0.706 | 0.874 | 1.000 | | | | | | | | | | | |
| (5, 5) | 0.008 | 0.040 | 0.167 | 0.357 | 0.643 | 0.833 | 0.960 | 0.992 | 1.000 | | | | | | | | | | |
| (5, 6) | 0.004 | 0.024 | 0.110 | 0.262 | 0.522 | 0.738 | 0.911 | 0.976 | 0.998 | 1.000 | | | | | | | | | |
| (5, 7) | 0.003 | 0.015 | 0.076 | 0.197 | 0.424 | 0.652 | 0.854 | 0.955 | 0.992 | 1.000 | | | | | | | | | |
| (5, 8) | 0.002 | 0.010 | 0.054 | 0.152 | 0.347 | 0.576 | 0.793 | 0.929 | 0.984 | 1.000 | | | | | | | | | |
| (5, 9) | 0.001 | 0.007 | 0.039 | 0.119 | 0.287 | 0.510 | 0.734 | 0.902 | 0.972 | 1.000 | | | | | | | | | |
| (5, 10) | 0.001 | 0.005 | 0.029 | 0.095 | 0.239 | 0.455 | 0.678 | 0.874 | 0.958 | 1.000 | | | | | | | | | |
| (6, 6) | 0.002 | 0.013 | 0.067 | 0.175 | 0.392 | 0.608 | 0.825 | 0.933 | 0.987 | 0.998 | 1.000 | | | | | | | | |
| (6, 7) | 0.001 | 0.008 | 0.043 | 0.121 | 0.296 | 0.500 | 0.733 | 0.879 | 0.966 | 0.992 | 0.999 | 1.000 | | | | | | | |
| (6, 8) | 0.001 | 0.005 | 0.028 | 0.086 | 0.226 | 0.413 | 0.646 | 0.821 | 0.937 | 0.984 | 0.998 | 1.000 | | | | | | | |
| (6, 9) | 0.000 | 0.003 | 0.019 | 0.063 | 0.175 | 0.343 | 0.566 | 0.762 | 0.902 | 0.972 | 0.994 | 1.000 | | | | | | | |
| (6, 10) | 0.000 | 0.002 | 0.013 | 0.047 | 0.137 | 0.288 | 0.497 | 0.706 | 0.864 | 0.958 | 0.990 | 1.000 | | | | | | | |
| (7, 7) | 0.001 | 0.004 | 0.025 | 0.078 | 0.209 | 0.383 | 0.617 | 0.791 | 0.922 | 0.975 | 0.996 | 0.999 | 1.000 | | | | | | |
| (7, 8) | 0.000 | 0.002 | 0.015 | 0.051 | 0.149 | 0.296 | 0.514 | 0.704 | 0.867 | 0.949 | 0.988 | 0.998 | 1.000 | 1.000 | | | | | |
| (7, 9) | 0.000 | 0.001 | 0.010 | 0.035 | 0.108 | 0.231 | 0.427 | 0.622 | 0.806 | 0.916 | 0.975 | 0.994 | 0.999 | 1.000 | | | | | |
| (7, 10) | 0.000 | 0.001 | 0.006 | 0.024 | 0.080 | 0.182 | 0.355 | 0.549 | 0.743 | 0.879 | 0.957 | 0.990 | 0.998 | 1.000 | | | | | |
| (8, 8) | 0.000 | 0.001 | 0.009 | 0.032 | 0.100 | 0.214 | 0.405 | 0.595 | 0.786 | 0.900 | 0.968 | 0.991 | 0.999 | 1.000 | 1.000 | | | | |
| (8, 9) | 0.000 | 0.001 | 0.005 | 0.022 | 0.069 | 0.157 | 0.319 | 0.500 | 0.702 | 0.843 | 0.939 | 0.980 | 0.996 | 0.999 | 1.000 | 1.000 | | | |
| (8, 10) | 0.000 | 0.000 | 0.003 | 0.013 | 0.048 | 0.117 | 0.251 | 0.419 | 0.621 | 0.782 | 0.903 | 0.964 | 0.990 | 0.998 | 1.000 | 1.000 | | | |
| (9, 9) | 0.000 | 0.000 | 0.003 | 0.012 | 0.044 | 0.109 | 0.238 | 0.399 | 0.601 | 0.762 | 0.891 | 0.956 | 0.988 | 0.997 | 1.000 | 1.000 | 1.000 | | |
| (9, 10) | 0.000 | 0.000 | 0.002 | 0.008 | 0.029 | 0.077 | 0.179 | 0.319 | 0.510 | 0.681 | 0.834 | 0.923 | 0.974 | 0.992 | 0.999 | 1.000 | 1.000 | 1.000 | |
| (10, 10) | 0.000 | 0.000 | 0.001 | 0.004 | 0.019 | 0.051 | 0.128 | 0.242 | 0.414 | 0.586 | 0.758 | 0.872 | 0.949 | 0.981 | 0.996 | 0.999 | 1.000 | 1.000 | 1.000 |

*Swed, Frieda S. (now Frieda S. Cohn) and Churchill Eisenhart, 1948, *Annals of Mathematical Statistics*, 14: 70-82.

Reprinted with the permission of the Institute of Mathematical Statistics and the authors.

†Although probabilities are given for (n_1, n_2) with $n_1 \leq n_2$, the table is symmetric, and can be used for (n_1, n_2) with

TABLE A12
Quantiles of the Wilcoxon Signed Ranks Test Statistic†‡

| | $W_{0.005}$ | $W_{0.01}$ | $W_{0.025}$ | $W_{0.05}$ | $W_{0.10}$ | $W_{0.20}$ | $W_{0.30}$ | $W_{0.40}$ | $W_{0.50}$ | $\frac{n(n+1)}{2}$ |
|-------|-------------|------------|-------------|------------|------------|------------|------------|------------|------------|--------------------|
| n = 4 | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 4 | 5 | 10 |
| 5 | 0 | 0 | 0 | 1 | 3 | 4 | 5 | 6 | 7.5 | 15 |
| 6 | 0 | 0 | 1 | 3 | 4 | 6 | 8 | 9 | 10.5 | 21 |
| 7 | 0 | 1 | 3 | 4 | 6 | 9 | 11 | 12 | 14 | 28 |
| 8 | 1 | 2 | 4 | 6 | 9 | 12 | 14 | 16 | 18 | 36 |
| 9 | 2 | 4 | 6 | 9 | 11 | 15 | 18 | 20 | 22.5 | 45 |
| 10 | 4 | 6 | 9 | 11 | 15 | 19 | 22 | 25 | 27.5 | 55 |
| 11 | 6 | 8 | 11 | 14 | 18 | 23 | 27 | 30 | 33 | 66 |
| 12 | 8 | 10 | 14 | 18 | 22 | 28 | 32 | 36 | 39 | 78 |
| 13 | 10 | 13 | 18 | 22 | 27 | 33 | 38 | 42 | 45.5 | 91 |
| 14 | 13 | 16 | 22 | 26 | 32 | 39 | 44 | 48 | 52.5 | 105 |
| 15 | 16 | 20 | 26 | 31 | 37 | 45 | 51 | 55 | 60 | 120 |
| 16 | 20 | 24 | 30 | 36 | 43 | 51 | 58 | 63 | 68 | 136 |
| 17 | 24 | 28 | 35 | 42 | 49 | 58 | 65 | 71 | 76.5 | 153 |
| 18 | 28 | 33 | 41 | 48 | 56 | 66 | 73 | 80 | 85.5 | 171 |
| 19 | 33 | 38 | 47 | 54 | 63 | 74 | 82 | 89 | 95 | 190 |
| 20 | 38 | 44 | 53 | 61 | 70 | 83 | 91 | 98 | 105 | 210 |
| 21 | 44 | 50 | 59 | 68 | 78 | 91 | 100 | 108 | 115.5 | 231 |
| 22 | 49 | 56 | 67 | 76 | 87 | 100 | 110 | 119 | 126.5 | 253 |
| 23 | 55 | 63 | 74 | 84 | 95 | 110 | 120 | 130 | 138 | 276 |
| 24 | 62 | 70 | 82 | 92 | 105 | 120 | 131 | 141 | 150 | 300 |
| 25 | 69 | 77 | 90 | 101 | 114 | 131 | 143 | 153 | 162.5 | 325 |
| 26 | 76 | 85 | 99 | 111 | 125 | 142 | 155 | 165 | 175.5 | 351 |
| 27 | 84 | 94 | 108 | 120 | 135 | 154 | 167 | 178 | 189 | 378 |
| 28 | 92 | 102 | 117 | 131 | 146 | 166 | 180 | 192 | 203 | 406 |
| 29 | 101 | 111 | 127 | 141 | 158 | 178 | 193 | 206 | 217.5 | 435 |
| 30 | 110 | 121 | 138 | 152 | 170 | 191 | 207 | 220 | 232.5 | 465 |
| 31 | 119 | 131 | 148 | 164 | 182 | 205 | 221 | 235 | 248 | 496 |
| 32 | 129 | 141 | 160 | 176 | 195 | 219 | 236 | 250 | 264 | 528 |
| 33 | 139 | 152 | 171 | 188 | 208 | 233 | 251 | 266 | 280.5 | 561 |
| 34 | 149 | 163 | 183 | 201 | 222 | 248 | 266 | 282 | 297.5 | 595 |
| 35 | 160 | 175 | 196 | 214 | 236 | 263 | 283 | 299 | 315 | 630 |
| 36 | 172 | 187 | 209 | 228 | 251 | 279 | 299 | 317 | 333 | 666 |
| 37 | 184 | 199 | 222 | 242 | 266 | 295 | 316 | 335 | 351.5 | 703 |
| 38 | 196 | 212 | 236 | 257 | 282 | 312 | 334 | 353 | 370.5 | 741 |
| 39 | 208 | 225 | 250 | 272 | 298 | 329 | 352 | 372 | 390 | 780 |
| 40 | 221 | 239 | 265 | 287 | 314 | 347 | 371 | 391 | 410 | 820 |
| 41 | 235 | 253 | 280 | 303 | 331 | 365 | 390 | 411 | 430.5 | 861 |
| 42 | 248 | 267 | 295 | 320 | 349 | 384 | 409 | 431 | 451.5 | 903 |
| 43 | 263 | 282 | 311 | 337 | 366 | 403 | 429 | 452 | 473 | 946 |
| 44 | 277 | 297 | 328 | 354 | 385 | 422 | 450 | 473 | 495 | 990 |
| 45 | 292 | 313 | 344 | 372 | 403 | 442 | 471 | 495 | 517.5 | 1035 |
| 46 | 308 | 329 | 362 | 390 | 423 | 463 | 492 | 517 | 540.5 | 1081 |
| 47 | 324 | 346 | 379 | 408 | 442 | 484 | 514 | 540 | 564 | 1128 |

(Table A12 continued on next page.)

TABLE A12 (Cont'd)

| | $w_{0.005}$ | $w_{0.01}$ | $w_{0.025}$ | $w_{0.05}$ | $w_{0.10}$ | $w_{0.20}$ | $w_{0.30}$ | $w_{0.40}$ | $w_{0.50}$ | $\frac{n(n+1)}{2}$ |
|-----|-------------|------------|-------------|------------|------------|------------|------------|------------|------------|--------------------|
| 48 | 340 | 363 | 397 | 428 | 463 | 505 | 536 | 563 | 588 | 1176 |
| 49 | 357 | 381 | 416 | 447 | 483 | 527 | 559 | 587 | 612.5 | 1225 |
| 50§ | 374 | 398 | 435 | 467 | 504 | 550 | 583 | 611 | 637.5 | 1275 |

†Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from H. L. Harter and D. B. Owen, 1970, *Selected Tables in Mathematical Statistics, Vol. 1.*, Markham Publishing Co., Chicago. The values in the Selected Tables had been developed and distributed by the Lederle Laboratories Division of American Cyanamid Co., in cooperation with the Department of Statistics, Florida State University, Tallahassee, Florida, and copyrighted in 1963 by American Cyanamid Co. and Florida State University. Reprinted here with the permission of John Wiley & Sons, Inc., W. J. Conover, Florida State University, and the American Cyanamid Co.

‡The entries in this table are quantiles w_p of the Wilcoxon signed ranks test statistic W^* given by Equation 9.8 for selected values of $p \leq .50$. Quantiles w_p for $p > .50$ may be computed from the equation

$$w_p = n(n+1)/2 - w_{1-p}$$

where $n(n+1)/2$ is given in the right hand column in the table. Note that $P(W^* < w_p) \leq p$ and $P(W^* > w_p) \leq 1 - p$ if H_0 is true. Critical regions correspond to values of W^* less than (or greater than) but not including the appropriate quantile.

§For n larger than 50, the p^{th} quantile w_p of the Wilcoxon signed ranks test statistic may be approximated by $w_p = [n(n+1)/4] + x_p \sqrt{n(n+1)(2n+1)/24}$, where x_p is the p^{th} quantile of a standard normal random variable, obtained from Table A3.

TABLE A13

Quantiles w_p of the Mann-Whitney Test Statistic*†

| n | p | m=2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20‡ | |
|---|------|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|
| 2 | .001 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| | .005 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 |
| | .01 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| | .025 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 |
| | .05 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 |
| | .10 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 |
| 3 | .001 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 7 |
| | .005 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 |
| | .01 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 11 | 12 |
| | .025 | 6 | 6 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 14 | 14 | 14 | 15 |
| | .05 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 11 | 11 | 12 | 12 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 17 | 18 |
| | .10 | 7 | 8 | 8 | 9 | 10 | 11 | 12 | 12 | 13 | 14 | 15 | 16 | 17 | 17 | 18 | 19 | 20 | 20 | 21 | 22 |
| 4 | .001 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 12 | 13 | 13 | 14 | 14 | 14 | 14 |
| | .005 | 10 | 10 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 17 | 18 | 18 | 19 |
| | .01 | 10 | 10 | 10 | 11 | 12 | 12 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 18 | 18 | 19 | 20 | 20 | 21 | 21 |
| | .025 | 10 | 10 | 11 | 12 | 13 | 14 | 15 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 22 | 23 | 24 | 24 | 25 |
| | .05 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 25 | 26 | 27 | 27 | 28 | 29 |
| | .10 | 11 | 12 | 14 | 15 | 16 | 17 | 18 | 20 | 21 | 22 | 23 | 24 | 26 | 27 | 28 | 29 | 31 | 32 | 32 | 33 |
| 5 | .001 | 15 | 15 | 15 | 15 | 15 | 15 | 16 | 17 | 17 | 18 | 18 | 19 | 19 | 20 | 21 | 21 | 22 | 23 | 23 | |
| | .005 | 15 | 15 | 15 | 16 | 17 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 23 | 24 | 25 | 26 | 27 | 28 | 28 | 29 |
| | .01 | 15 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 31 | 32 |
| | .025 | 15 | 16 | 17 | 18 | 19 | 21 | 22 | 23 | 24 | 25 | 27 | 28 | 29 | 30 | 31 | 33 | 34 | 35 | 35 | 36 |
| | .05 | 16 | 17 | 18 | 20 | 21 | 22 | 24 | 25 | 27 | 28 | 29 | 31 | 32 | 34 | 35 | 36 | 38 | 39 | 39 | 41 |
| | .10 | 17 | 18 | 20 | 21 | 23 | 24 | 26 | 28 | 29 | 31 | 33 | 34 | 36 | 38 | 39 | 41 | 43 | 44 | 44 | 46 |
| 6 | .001 | 21 | 21 | 21 | 21 | 21 | 21 | 23 | 24 | 25 | 26 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | |
| | .005 | 21 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 31 | 32 | 33 | 34 | 35 | 37 | 38 | 39 | 39 | 40 |
| | .01 | 21 | 21 | 23 | 24 | 25 | 26 | 28 | 29 | 30 | 31 | 33 | 34 | 35 | 37 | 38 | 40 | 41 | 42 | 42 | 44 |
| | .025 | 21 | 23 | 24 | 25 | 27 | 28 | 30 | 32 | 33 | 35 | 36 | 38 | 39 | 41 | 43 | 44 | 46 | 47 | 47 | 49 |
| | .05 | 22 | 24 | 25 | 27 | 29 | 30 | 32 | 34 | 36 | 38 | 39 | 41 | 43 | 45 | 47 | 48 | 50 | 52 | 52 | 54 |
| | .10 | 23 | 25 | 27 | 29 | 31 | 33 | 35 | 37 | 39 | 41 | 43 | 45 | 47 | 49 | 51 | 53 | 56 | 58 | 58 | 60 |

(Table A13 continued on next page.)

TABLE A13 (Cont'd)

| n | p | m=2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20† |
|----|------|-----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|
| 7 | .001 | 28 | 28 | 28 | 28 | 29 | 30 | 31 | 32 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 42 | 43 | 44 | 45 |
| | .005 | 28 | 28 | 29 | 30 | 32 | 33 | 35 | 36 | 38 | 39 | 41 | 42 | 44 | 45 | 47 | 48 | 50 | 51 | 53 |
| | .01 | 28 | 29 | 30 | 32 | 33 | 35 | 36 | 38 | 40 | 41 | 43 | 45 | 46 | 48 | 50 | 52 | 53 | 55 | 57 |
| | .025 | 28 | 30 | 32 | 34 | 35 | 37 | 39 | 41 | 43 | 45 | 47 | 49 | 51 | 53 | 55 | 57 | 59 | 61 | 63 |
| | .05 | 29 | 31 | 33 | 35 | 37 | 40 | 42 | 44 | 46 | 48 | 50 | 53 | 55 | 57 | 59 | 62 | 64 | 66 | 68 |
| | .10 | 30 | 33 | 35 | 37 | 40 | 42 | 45 | 47 | 50 | 52 | 55 | 57 | 60 | 62 | 65 | 67 | 70 | 72 | 75 |
| 8 | .001 | 36 | 36 | 36 | 37 | 38 | 39 | 41 | 42 | 43 | 45 | 46 | 48 | 49 | 51 | 52 | 54 | 55 | 57 | 58 |
| | .005 | 36 | 36 | 38 | 39 | 41 | 43 | 44 | 46 | 48 | 50 | 52 | 54 | 55 | 57 | 59 | 61 | 63 | 65 | 67 |
| | .01 | 36 | 37 | 39 | 41 | 43 | 44 | 46 | 48 | 50 | 52 | 54 | 56 | 59 | 61 | 63 | 65 | 67 | 69 | 71 |
| | .025 | 37 | 39 | 41 | 43 | 45 | 47 | 50 | 52 | 54 | 56 | 59 | 61 | 63 | 66 | 68 | 71 | 73 | 75 | 78 |
| | .05 | 38 | 40 | 42 | 45 | 47 | 50 | 52 | 55 | 57 | 60 | 63 | 65 | 68 | 70 | 73 | 76 | 78 | 81 | 84 |
| | .10 | 39 | 42 | 44 | 47 | 50 | 53 | 56 | 59 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 88 | 91 |
| 9 | .001 | 45 | 45 | 45 | 47 | 48 | 49 | 51 | 53 | 54 | 56 | 58 | 60 | 61 | 63 | 65 | 67 | 69 | 71 | 72 |
| | .005 | 45 | 46 | 47 | 49 | 51 | 53 | 55 | 57 | 59 | 62 | 64 | 66 | 68 | 70 | 73 | 75 | 77 | 79 | 82 |
| | .01 | 45 | 47 | 49 | 51 | 53 | 55 | 57 | 60 | 62 | 64 | 67 | 69 | 72 | 74 | 77 | 79 | 82 | 84 | 86 |
| | .025 | 46 | 48 | 50 | 53 | 56 | 58 | 61 | 63 | 66 | 69 | 72 | 74 | 77 | 80 | 83 | 85 | 88 | 91 | 94 |
| | .05 | 47 | 50 | 52 | 55 | 58 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 88 | 91 | 94 | 97 | 100 |
| | .10 | 48 | 51 | 55 | 58 | 61 | 64 | 68 | 71 | 74 | 77 | 81 | 84 | 87 | 91 | 94 | 98 | 101 | 104 | 108 |
| 10 | .001 | 55 | 55 | 56 | 57 | 59 | 61 | 62 | 64 | 66 | 68 | 70 | 73 | 75 | 77 | 79 | 81 | 83 | 85 | 88 |
| | .005 | 55 | 56 | 58 | 60 | 62 | 65 | 67 | 69 | 72 | 74 | 77 | 80 | 82 | 85 | 87 | 90 | 93 | 95 | 98 |
| | .01 | 55 | 57 | 59 | 62 | 64 | 67 | 69 | 72 | 75 | 78 | 80 | 83 | 86 | 89 | 92 | 94 | 97 | 100 | 103 |
| | .025 | 56 | 59 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 89 | 92 | 95 | 98 | 101 | 104 | 108 | 111 |
| | .05 | 57 | 60 | 63 | 67 | 70 | 73 | 76 | 80 | 83 | 87 | 90 | 93 | 97 | 100 | 104 | 107 | 111 | 114 | 118 |
| | .10 | 59 | 62 | 66 | 69 | 73 | 77 | 80 | 84 | 88 | 92 | 95 | 99 | 103 | 107 | 110 | 114 | 118 | 122 | 126 |

| n | p | m=2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20† |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11 | .001 | 66 | 66 | 67 | 69 | 71 | 73 | 75 | 77 | 79 | 82 | 84 | 87 | 89 | 91 | 94 | 96 | 99 | 101 | 104 |
| | .005 | 66 | 67 | 69 | 72 | 74 | 77 | 80 | 83 | 85 | 88 | 91 | 94 | 97 | 100 | 103 | 106 | 109 | 112 | 115 |
| | .01 | 66 | 68 | 71 | 74 | 76 | 79 | 82 | 85 | 89 | 92 | 95 | 98 | 101 | 104 | 108 | 111 | 114 | 117 | 120 |
| | .025 | 67 | 70 | 73 | 76 | 80 | 83 | 86 | 90 | 93 | 97 | 100 | 104 | 107 | 111 | 114 | 118 | 122 | 125 | 129 |
| | .05 | 68 | 72 | 75 | 79 | 83 | 86 | 90 | 94 | 98 | 101 | 105 | 109 | 113 | 117 | 121 | 124 | 128 | 132 | 136 |
| | .10 | 70 | 74 | 78 | 82 | 86 | 90 | 94 | 98 | 98 | 103 | 107 | 111 | 115 | 119 | 124 | 128 | 132 | 136 | 140 |
| 12 | .001 | 78 | 78 | 79 | 81 | 83 | 86 | 88 | 91 | 93 | 96 | 98 | 102 | 104 | 106 | 110 | 113 | 116 | 118 | 121 |
| | .005 | 78 | 80 | 82 | 85 | 88 | 91 | 94 | 97 | 100 | 103 | 106 | 110 | 113 | 116 | 120 | 123 | 126 | 130 | 133 |
| | .01 | 78 | 81 | 84 | 87 | 90 | 93 | 96 | 100 | 103 | 107 | 110 | 114 | 117 | 121 | 125 | 128 | 132 | 135 | 139 |
| | .025 | 80 | 83 | 86 | 90 | 93 | 97 | 101 | 105 | 108 | 112 | 116 | 120 | 124 | 128 | 132 | 136 | 140 | 144 | 148 |
| | .05 | 81 | 84 | 88 | 92 | 96 | 100 | 105 | 109 | 111 | 117 | 121 | 126 | 130 | 134 | 139 | 143 | 147 | 151 | 156 |
| | .10 | 83 | 87 | 91 | 96 | 100 | 105 | 109 | 114 | 118 | 123 | 128 | 132 | 137 | 142 | 146 | 151 | 156 | 160 | 165 |
| 13 | .001 | 91 | 91 | 93 | 95 | 97 | 100 | 103 | 106 | 109 | 112 | 115 | 118 | 121 | 124 | 127 | 130 | 134 | 137 | 140 |
| | .005 | 91 | 93 | 95 | 99 | 102 | 105 | 109 | 112 | 116 | 119 | 123 | 126 | 130 | 134 | 137 | 141 | 145 | 149 | 152 |
| | .01 | 92 | 94 | 97 | 101 | 104 | 108 | 112 | 115 | 119 | 123 | 127 | 131 | 135 | 139 | 143 | 147 | 151 | 155 | 159 |
| | .025 | 93 | 96 | 100 | 104 | 108 | 112 | 116 | 120 | 125 | 129 | 133 | 137 | 142 | 146 | 151 | 155 | 159 | 164 | 168 |
| | .05 | 94 | 98 | 102 | 107 | 111 | 116 | 120 | 125 | 129 | 134 | 139 | 143 | 148 | 153 | 157 | 162 | 167 | 172 | 176 |
| | .10 | 96 | 101 | 105 | 110 | 115 | 120 | 125 | 130 | 135 | 140 | 145 | 150 | 155 | 160 | 166 | 171 | 176 | 181 | 186 |
| 14 | .001 | 105 | 105 | 107 | 109 | 112 | 115 | 118 | 121 | 125 | 128 | 131 | 135 | 138 | 142 | 145 | 149 | 152 | 156 | 160 |
| | .005 | 105 | 107 | 110 | 113 | 117 | 121 | 124 | 128 | 132 | 136 | 140 | 144 | 148 | 152 | 156 | 160 | 164 | 169 | 173 |
| | .01 | 106 | 108 | 112 | 116 | 119 | 123 | 128 | 132 | 136 | 140 | 144 | 149 | 153 | 157 | 162 | 166 | 171 | 175 | 179 |
| | .025 | 107 | 111 | 115 | 119 | 123 | 128 | 132 | 137 | 142 | 146 | 151 | 156 | 161 | 165 | 170 | 175 | 180 | 184 | 189 |
| | .05 | 109 | 113 | 117 | 122 | 127 | 132 | 137 | 142 | 147 | 152 | 157 | 162 | 167 | 172 | 177 | 183 | 188 | 193 | 198 |
| | .10 | 110 | 116 | 121 | 126 | 131 | 137 | 142 | 147 | 153 | 158 | 164 | 169 | 175 | 180 | 186 | 191 | 197 | 203 | 208 |

(Table A13 continued on next page.)

TABLE A13 (Cont'd)

| n | p | m=2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20† |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 15 | .001 | 120 | 120 | 122 | 125 | 128 | 133 | 135 | 138 | 142 | 145 | 149 | 153 | 157 | 161 | 164 | 168 | 172 | 176 | 180 |
| | .005 | 120 | 123 | 126 | 129 | 133 | 137 | 141 | 145 | 150 | 154 | 158 | 163 | 167 | 172 | 176 | 181 | 185 | 190 | 194 |
| | .01 | 121 | 124 | 128 | 132 | 136 | 140 | 145 | 149 | 154 | 158 | 163 | 168 | 172 | 177 | 182 | 187 | 191 | 196 | 201 |
| | .025 | 122 | 126 | 131 | 135 | 140 | 145 | 150 | 155 | 160 | 165 | 170 | 175 | 180 | 185 | 191 | 196 | 201 | 206 | 211 |
| | .05 | 124 | 128 | 133 | 139 | 144 | 149 | 154 | 160 | 165 | 171 | 176 | 182 | 187 | 193 | 198 | 204 | 209 | 215 | 221 |
| | .10 | 126 | 131 | 137 | 143 | 148 | 154 | 160 | 166 | 172 | 178 | 184 | 189 | 195 | 201 | 207 | 213 | 219 | 225 | 231 |
| 16 | .001 | 136 | 136 | 139 | 142 | 145 | 148 | 152 | 156 | 160 | 164 | 168 | 172 | 176 | 180 | 185 | 189 | 193 | 197 | 202 |
| | .005 | 136 | 139 | 142 | 146 | 150 | 155 | 159 | 164 | 168 | 173 | 178 | 182 | 187 | 192 | 197 | 202 | 207 | 211 | 216 |
| | .01 | 137 | 140 | 144 | 149 | 153 | 158 | 163 | 168 | 173 | 178 | 183 | 188 | 193 | 198 | 203 | 208 | 213 | 219 | 224 |
| | .025 | 138 | 143 | 148 | 152 | 158 | 163 | 168 | 174 | 179 | 184 | 190 | 196 | 201 | 207 | 212 | 218 | 223 | 229 | 235 |
| | .05 | 140 | 145 | 151 | 156 | 162 | 167 | 173 | 179 | 185 | 191 | 197 | 202 | 208 | 214 | 220 | 226 | 232 | 238 | 244 |
| | .10 | 142 | 148 | 154 | 160 | 166 | 173 | 179 | 185 | 191 | 198 | 204 | 211 | 217 | 223 | 230 | 236 | 243 | 249 | 256 |
| 17 | .001 | 153 | 154 | 156 | 159 | 163 | 167 | 171 | 175 | 179 | 183 | 188 | 192 | 197 | 201 | 206 | 211 | 215 | 220 | 224 |
| | .005 | 153 | 156 | 160 | 164 | 169 | 173 | 178 | 183 | 188 | 193 | 198 | 203 | 208 | 214 | 219 | 224 | 229 | 235 | 240 |
| | .01 | 154 | 158 | 162 | 167 | 172 | 177 | 182 | 187 | 192 | 198 | 203 | 209 | 214 | 220 | 225 | 231 | 236 | 242 | 247 |
| | .025 | 156 | 160 | 165 | 171 | 176 | 182 | 188 | 193 | 199 | 205 | 211 | 217 | 223 | 229 | 235 | 241 | 247 | 253 | 259 |
| | .05 | 157 | 163 | 169 | 174 | 180 | 187 | 193 | 199 | 205 | 211 | 218 | 224 | 231 | 237 | 243 | 250 | 256 | 263 | 269 |
| | .10 | 160 | 166 | 172 | 179 | 185 | 192 | 199 | 206 | 212 | 219 | 226 | 233 | 239 | 246 | 253 | 260 | 267 | 274 | 281 |
| 18 | .001 | 171 | 172 | 175 | 178 | 182 | 186 | 190 | 195 | 199 | 204 | 209 | 214 | 218 | 223 | 228 | 233 | 238 | 243 | 248 |
| | .005 | 171 | 174 | 178 | 183 | 188 | 193 | 198 | 203 | 209 | 214 | 219 | 225 | 230 | 236 | 242 | 247 | 253 | 259 | 264 |
| | .01 | 172 | 176 | 181 | 186 | 191 | 196 | 202 | 208 | 213 | 219 | 225 | 231 | 237 | 242 | 248 | 254 | 260 | 266 | 272 |
| | .025 | 174 | 179 | 184 | 190 | 196 | 202 | 208 | 214 | 220 | 227 | 233 | 239 | 246 | 252 | 258 | 265 | 271 | 278 | 284 |
| | .05 | 176 | 181 | 188 | 194 | 200 | 207 | 213 | 220 | 227 | 233 | 240 | 247 | 254 | 260 | 267 | 274 | 281 | 288 | 295 |
| | .10 | 178 | 185 | 192 | 199 | 206 | 213 | 220 | 227 | 234 | 241 | 249 | 256 | 263 | 270 | 278 | 285 | 292 | 300 | 307 |
| 19 | .001 | 190 | 191 | 194 | 198 | 202 | 206 | 211 | 216 | 220 | 225 | 231 | 236 | 241 | 246 | 251 | 257 | 262 | 268 | 273 |
| | .005 | 191 | 194 | 198 | 203 | 208 | 213 | 219 | 224 | 230 | 236 | 242 | 248 | 254 | 260 | 265 | 272 | 278 | 284 | 290 |
| | .01 | 192 | 195 | 200 | 206 | 211 | 217 | 223 | 229 | 235 | 241 | 247 | 254 | 260 | 266 | 273 | 279 | 285 | 292 | 298 |
| | .025 | 193 | 198 | 204 | 210 | 216 | 223 | 229 | 236 | 243 | 249 | 256 | 263 | 269 | 276 | 283 | 290 | 297 | 304 | 310 |
| | .05 | 195 | 201 | 208 | 214 | 221 | 228 | 235 | 242 | 249 | 256 | 263 | 271 | 278 | 285 | 292 | 300 | 307 | 314 | 321 |
| | .10 | 198 | 205 | 212 | 219 | 227 | 234 | 242 | 249 | 257 | 264 | 272 | 280 | 288 | 295 | 303 | 311 | 319 | 326 | 334 |

| n | p | m=2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20‡ |
|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 20‡ | .001 | 210 | 211 | 214 | 218 | 223 | 227 | 232 | 237 | 243 | 248 | 253 | 259 | 265 | 270 | 276 | 281 | 287 | 293 | 299 |
| | .005 | 211 | 214 | 219 | 224 | 229 | 235 | 241 | 247 | 253 | 259 | 265 | 271 | 278 | 284 | 290 | 297 | 303 | 310 | 316 |
| | .01 | 212 | 216 | 221 | 227 | 233 | 239 | 245 | 251 | 258 | 264 | 271 | 278 | 284 | 291 | 298 | 304 | 311 | 318 | 325 |
| | .025 | 213 | 219 | 225 | 231 | 238 | 245 | 251 | 259 | 266 | 273 | 280 | 287 | 294 | 301 | 309 | 316 | 323 | 330 | 338 |
| | .05 | 215 | 222 | 229 | 236 | 243 | 250 | 258 | 265 | 273 | 280 | 288 | 295 | 303 | 311 | 318 | 326 | 334 | 341 | 349 |
| | .10 | 218 | 226 | 233 | 241 | 249 | 257 | 265 | 273 | 281 | 289 | 297 | 305 | 313 | 321 | 330 | 338 | 346 | 354 | 362 |

*Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © 1980, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc., and the author.

†The entries in this table are quantiles w_p of the Mann-Whitney test statistic W' , given by Equation 9.10 for selected values of p . Note that $\Pr(W' < w_p) \leq p$. Upper quantiles may be found from the equation

$$w_p = n(n + m + 1) - w_{1-p}$$

Critical regions correspond to values of W' less than (or greater than) but not equal to the appropriate quantile.

‡For n or m greater than 20, the p^{th} quantile w_p of the Mann-Whitney test statistic may be approximated by

$$w_p = n(N + 1)/2 + x_p \sqrt{nm(N + 1)/12}$$

where x_p is the p^{th} quantile of a standard normal random variable, obtained from Table A3, and where $N = m + n$.

TABLE A14
Quantiles of the Squared Ranks Test Statistic†‡

| n | p | m=3 | 4 | 5 | 6 | 7 | 8 | 9 | 10§ |
|----------|----------|------------|----------|----------|----------|----------|----------|----------|------------|
| 3 | .005 | 14 | 14 | 14 | 14 | 14 | 14 | 21 | 21 |
| | .01 | 14 | 14 | 14 | 14 | 21 | 21 | 26 | 26 |
| | .025 | 14 | 14 | 21 | 26 | 29 | 30 | 35 | 41 |
| | .05 | 21 | 21 | 26 | 30 | 38 | 42 | 49 | 54 |
| | .10 | 26 | 29 | 35 | 42 | 50 | 59 | 69 | 77 |
| | .90 | 65 | 90 | 117 | 149 | 182 | 221 | 260 | 305 |
| | .95 | 70 | 101 | 129 | 161 | 197 | 238 | 285 | 333 |
| | .975 | 77 | 110 | 138 | 170 | 213 | 257 | 308 | 362 |
| | .99 | 77 | 110 | 149 | 194 | 230 | 285 | 329 | 394 |
| | .995 | 77 | 110 | 149 | 194 | 245 | 302 | 346 | 413 |
| 4 | .005 | 30 | 30 | 30 | 39 | 39 | 46 | 50 | 54 |
| | .01 | 30 | 30 | 39 | 46 | 50 | 51 | 62 | 66 |
| | .025 | 30 | 39 | 50 | 54 | 63 | 71 | 78 | 90 |
| | .05 | 39 | 50 | 57 | 66 | 78 | 90 | 102 | 114 |
| | .10 | 50 | 62 | 71 | 85 | 99 | 114 | 130 | 149 |
| | .90 | 111 | 142 | 182 | 222 | 270 | 321 | 375 | 435 |
| | .95 | 119 | 154 | 197 | 246 | 294 | 350 | 413 | 476 |
| | .975 | 126 | 165 | 206 | 255 | 311 | 374 | 439 | 510 |
| | .99 | 126 | 174 | 219 | 270 | 334 | 401 | 470 | 545 |
| | .995 | 126 | 174 | 230 | 281 | 351 | 414 | 494 | 567 |
| 5 | .005 | 55 | 55 | 66 | 75 | 79 | 88 | 99 | 110 |
| | .01 | 55 | 66 | 75 | 82 | 90 | 103 | 115 | 127 |
| | .025 | 66 | 79 | 88 | 100 | 114 | 130 | 145 | 162 |
| | .05 | 75 | 88 | 103 | 120 | 135 | 155 | 175 | 195 |
| | .10 | 87 | 103 | 121 | 142 | 163 | 187 | 212 | 239 |
| | .90 | 169 | 214 | 264 | 319 | 379 | 445 | 514 | 591 |
| | .95 | 178 | 228 | 282 | 342 | 410 | 479 | 558 | 639 |
| | .975 | 183 | 235 | 297 | 363 | 433 | 508 | 592 | 680 |
| | .99 | 190 | 246 | 310 | 382 | 459 | 543 | 631 | 727 |
| | .995 | 190 | 255 | 319 | 391 | 478 | 559 | 654 | 754 |
| 6 | .005 | 91 | 104 | 115 | 124 | 136 | 152 | 167 | 182 |
| | .01 | 91 | 115 | 124 | 139 | 155 | 175 | 191 | 210 |
| | .025 | 115 | 130 | 143 | 164 | 184 | 208 | 231 | 255 |
| | .05 | 124 | 139 | 164 | 187 | 211 | 239 | 268 | 299 |
| | .10 | 136 | 163 | 187 | 215 | 247 | 280 | 315 | 352 |
| | .90 | 243 | 300 | 364 | 435 | 511 | 592 | 679 | 772 |
| | .95 | 255 | 319 | 386 | 463 | 545 | 634 | 730 | 831 |
| | .975 | 259 | 331 | 406 | 486 | 574 | 670 | 771 | 880 |
| | .99 | 271 | 339 | 424 | 511 | 607 | 706 | 817 | 935 |
| | .995 | 271 | 346 | 431 | 526 | 624 | 731 | 847 | 970 |
| 7 | .005 | 140 | 155 | 172 | 195 | 212 | 235 | 257 | 280 |
| | .01 | 155 | 172 | 191 | 212 | 236 | 260 | 287 | 315 |
| | .025 | 172 | 195 | 217 | 245 | 274 | 305 | 338 | 372 |
| | .05 | 188 | 212 | 240 | 274 | 308 | 344 | 384 | 425 |
| | .10 | 203 | 236 | 271 | 308 | 350 | 394 | 440 | 489 |
| | .90 | 335 | 407 | 487 | 572 | 665 | 764 | 871 | 984 |
| | .95 | 347 | 428 | 515 | 608 | 707 | 814 | 929 | 1051 |
| | .975 | 356 | 443 | 536 | 635 | 741 | 856 | 979 | 1108 |
| | .99 | 364 | 456 | 560 | 664 | 779 | 900 | 1032 | 1172 |
| | .995 | 371 | 467 | 571 | 683 | 803 | 929 | 1067 | 1212 |

(Table A14 continued on next page.)

TABLE A14 (Cont'd)

| n | p | m=3 | 4 | 5 | 6 | 7 | 8 | 9 | 10§ |
|-----|------|-----|-----|------|------|------|------|------|------|
| 8 | .005 | 204 | 236 | 260 | 284 | 311 | 340 | 368 | 401 |
| | .01 | 221 | 249 | 276 | 309 | 340 | 372 | 408 | 445 |
| | .025 | 249 | 276 | 311 | 345 | 384 | 425 | 468 | 513 |
| | .05 | 268 | 300 | 340 | 381 | 426 | 473 | 524 | 576 |
| | .10 | 285 | 329 | 374 | 423 | 476 | 531 | 590 | 652 |
| | .90 | 447 | 536 | 632 | 735 | 846 | 965 | 1091 | 1224 |
| | .95 | 464 | 560 | 664 | 776 | 896 | 1023 | 1159 | 1303 |
| | .975 | 476 | 579 | 689 | 807 | 935 | 1071 | 1215 | 1368 |
| | .99 | 485 | 599 | 716 | 840 | 980 | 1124 | 1277 | 1442 |
| | .995 | 492 | 604 | 731 | 863 | 1005 | 1156 | 1319 | 1489 |
| 9 | .005 | 304 | 325 | 361 | 393 | 429 | 466 | 508 | 549 |
| | .01 | 321 | 349 | 384 | 423 | 464 | 508 | 553 | 601 |
| | .025 | 342 | 380 | 423 | 469 | 517 | 570 | 624 | 682 |
| | .05 | 365 | 406 | 457 | 510 | 567 | 626 | 689 | 755 |
| | .10 | 390 | 444 | 501 | 561 | 625 | 694 | 766 | 843 |
| | .90 | 581 | 689 | 803 | 925 | 1056 | 1195 | 1343 | 1498 |
| | .95 | 601 | 717 | 840 | 972 | 1112 | 1251 | 1420 | 1587 |
| | .975 | 615 | 741 | 870 | 1009 | 1158 | 1317 | 1485 | 1662 |
| | .99 | 624 | 757 | 900 | 1049 | 1209 | 1377 | 1556 | 1745 |
| | .995 | 629 | 769 | 916 | 1073 | 1239 | 1417 | 1601 | 1798 |
| 10§ | .005 | 406 | 448 | 486 | 526 | 573 | 620 | 672 | 725 |
| | .01 | 425 | 470 | 513 | 561 | 613 | 667 | 725 | 785 |
| | .025 | 457 | 505 | 560 | 616 | 677 | 741 | 808 | 879 |
| | .05 | 486 | 539 | 601 | 665 | 734 | 806 | 883 | 963 |
| | .10 | 514 | 580 | 649 | 724 | 801 | 885 | 972 | 1064 |
| | .90 | 742 | 866 | 1001 | 1144 | 1296 | 1457 | 1627 | 1806 |
| | .95 | 765 | 901 | 1045 | 1197 | 1360 | 1533 | 1715 | 1907 |
| | .975 | 778 | 925 | 1078 | 1241 | 1413 | 1596 | 1788 | 1991 |
| | .99 | 793 | 949 | 1113 | 1286 | 1470 | 1664 | 1869 | 2085 |
| | .995 | 798 | 961 | 1130 | 1314 | 1505 | 1708 | 1921 | 2145 |

†Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from tables by Conover and Iman, 1978, *Communications in Statistics*, B7: 491-513, Marcel-Dekker Journals, New York. Reprinted with the permission of John Wiley & Sons, Inc., Marcel-Dekker, Inc., and authors W. J. Conover and R. L. Iman.

‡The entries in this table are selected quantiles w_p of the squared ranks test statistic T_x^* , given by Equation 9.12. Note that $\Pr(T_x^* < w_p) \leq p$ and $\Pr(T_x^* > w_p) \leq 1 - p$. Critical regions correspond to values less than (or greater than) but not including the appropriate quantile.

§For n or m greater than 10, the p^{th} quantile w_p of the squared ranks test statistic may be approximated by

$$w_p = \frac{n(N + 1)(2N + 1)}{6} + x_p \sqrt{\frac{mn(N + 1)(2N + 1)(8N + 11)}{180}}$$

where $N = n + m$, and x_p is the p^{th} quantile of a standard normal random variable, obtained from Table A3.

TABLE A15
Quantiles of the Spearman Test Statistic*†

| n | p = .900 | .950 | .975 | .990 | .995 | .999 |
|-----|----------|-------|-------|-------|-------|-------|
| 4 | .8000 | .8000 | | | | |
| 5 | .7000 | .8000 | .9000 | .9000 | | |
| 6 | .6000 | .7714 | .8286 | .8857 | .9429 | |
| 7 | .5357 | .6786 | .7450 | .8571 | .8929 | .9643 |
| 8 | .5000 | .6190 | .7143 | .8095 | .8571 | .9286 |
| 9 | .4667 | .5833 | .6833 | .7667 | .8167 | .9000 |
| 10 | .4424 | .5515 | .6364 | .7333 | .7818 | .8667 |
| 11 | .4182 | .5273 | .6091 | .7000 | .7455 | .8364 |
| 12 | .3986 | .4965 | .5804 | .6713 | .7273 | .8182 |
| 13 | .3791 | .4780 | .5549 | .6429 | .6978 | .7912 |
| 14 | .3626 | .4593 | .5341 | .6220 | .6747 | .7670 |
| 15 | .3500 | .4429 | .5179 | .6000 | .6536 | .7464 |
| 16 | .3382 | .4265 | .5000 | .5824 | .6324 | .7265 |
| 17 | .3260 | .4118 | .4853 | .5637 | .6152 | .7083 |
| 18 | .3148 | .3994 | .4716 | .5480 | .5975 | .6904 |
| 19 | .3070 | .3895 | .4579 | .5333 | .5825 | .6737 |
| 20 | .2977 | .3789 | .4451 | .5203 | .5684 | .6586 |
| 21 | .2909 | .3688 | .4351 | .5078 | .5545 | .6455 |
| 22 | .2829 | .3597 | .4241 | .4963 | .5426 | .6318 |
| 23 | .2767 | .3518 | .4150 | .4852 | .5306 | .6186 |
| 24 | .2704 | .3435 | .4061 | .4748 | .5200 | .6070 |
| 25 | .2646 | .3362 | .3977 | .4654 | .5100 | .5962 |
| 26 | .2588 | .3299 | .3894 | .4564 | .5002 | .5856 |
| 27 | .2540 | .3236 | .3822 | .4481 | .4915 | .5757 |
| 28 | .2490 | .3175 | .3749 | .4401 | .4828 | .5660 |
| 29 | .2443 | .3113 | .3685 | .4320 | .4744 | .5567 |
| 30‡ | .2400 | .3059 | .3620 | .4251 | .4665 | .5479 |

*Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © 1980, John Wiley & Sons, Inc. Adapted from G. J. Glasser and R. F. Winter, 1961, *Biometrika*, 48: 444. Reprinted with the permission of John Wiley & Sons, Inc., the Biometrika Trustees, and the principal authors, W. J. Conover and G. J. Glasser.

†The entries in this table are selected quantiles w_p of the Spearman rank correlation coefficient $\bar{\rho}$ when used as a test statistic. The lower quantiles may be obtained from the equation

$$w_p = -w_{1-p}$$

The critical region corresponds to values of $\bar{\rho}$ smaller than (or greater than) but not including the appropriate quantile. Note that the median of $\bar{\rho}$ is 0.

‡For n greater than 30 the approximate quantile of $\bar{\rho}$ may be obtained from

$$w_p \cong \frac{x_p}{\sqrt{n-1}}$$

where x_p is the p^{th} quantile of a standard normal random variable, obtained from Table A3.

TABLE A16

Quantiles of the Kendall Test Statistic†‡

| n | p=.900 | .950 | .975 | .990 | .995 |
|----|--------|------|------|------|------|
| 4 | 4 | 4 | 6 | 6 | 6 |
| 5 | 6 | 6 | 8 | 8 | 10 |
| 6 | 7 | 9 | 11 | 11 | 13 |
| 7 | 9 | 11 | 13 | 15 | 17 |
| 8 | 10 | 14 | 16 | 18 | 20 |
| 9 | 12 | 16 | 18 | 22 | 24 |
| 10 | 15 | 19 | 21 | 25 | 27 |
| 11 | 17 | 21 | 25 | 29 | 31 |
| 12 | 18 | 24 | 28 | 34 | 36 |
| 13 | 22 | 26 | 32 | 38 | 42 |
| 14 | 23 | 31 | 35 | 41 | 45 |
| 15 | 27 | 33 | 39 | 47 | 51 |
| 16 | 28 | 36 | 44 | 50 | 56 |
| 17 | 32 | 40 | 48 | 56 | 62 |
| 18 | 35 | 43 | 51 | 61 | 67 |
| 19 | 37 | 47 | 55 | 65 | 73 |
| 20 | 40 | 50 | 60 | 70 | 78 |
| 21 | 42 | 54 | 64 | 76 | 84 |
| 22 | 45 | 59 | 69 | 81 | 89 |
| 23 | 49 | 63 | 73 | 87 | 97 |
| 24 | 52 | 66 | 78 | 92 | 102 |
| 25 | 56 | 70 | 84 | 98 | 108 |
| 26 | 59 | 75 | 89 | 105 | 115 |
| 27 | 61 | 79 | 93 | 111 | 123 |
| 28 | 66 | 84 | 98 | 116 | 128 |
| 29 | 68 | 88 | 104 | 124 | 136 |
| 30 | 73 | 93 | 109 | 129 | 143 |
| 31 | 75 | 97 | 115 | 135 | 149 |
| 32 | 80 | 102 | 120 | 142 | 158 |
| 33 | 84 | 106 | 126 | 150 | 164 |
| 34 | 87 | 111 | 131 | 155 | 173 |
| 35 | 91 | 115 | 137 | 163 | 179 |
| 36 | 94 | 120 | 144 | 170 | 188 |
| 37 | 98 | 126 | 150 | 176 | 196 |
| 38 | 103 | 131 | 155 | 183 | 203 |
| 39 | 107 | 137 | 161 | 191 | 211 |
| 40 | 110 | 142 | 168 | 198 | 220 |
| 41 | 114 | 146 | 174 | 206 | 228 |
| 42 | 119 | 151 | 181 | 213 | 235 |
| 43 | 123 | 157 | 187 | 221 | 245 |
| 44 | 128 | 162 | 194 | 228 | 252 |
| 45 | 132 | 168 | 200 | 236 | 262 |
| 46 | 135 | 173 | 207 | 245 | 271 |
| 47 | 141 | 179 | 213 | 253 | 279 |
| 48 | 144 | 186 | 220 | 260 | 288 |
| 49 | 150 | 190 | 228 | 268 | 296 |

(Table A16 continued on next page.)

TABLE 16 (Cont'd)

| n | p = .900 | .950 | .975 | .990 | .995 |
|-----|----------|------|------|------|------|
| 50 | 153 | 197 | 233 | 277 | 305 |
| 51 | 159 | 203 | 241 | 285 | 315 |
| 52 | 162 | 208 | 248 | 294 | 324 |
| 53 | 168 | 214 | 256 | 302 | 334 |
| 54 | 173 | 221 | 263 | 311 | 343 |
| 55 | 177 | 227 | 269 | 319 | 353 |
| 56 | 182 | 232 | 276 | 328 | 362 |
| 57 | 186 | 240 | 284 | 336 | 372 |
| 58 | 191 | 245 | 291 | 345 | 381 |
| 59 | 197 | 251 | 299 | 355 | 391 |
| 60§ | 202 | 258 | 306 | 364 | 402 |

†Conover, W. J., 1980, *Practical Non-parametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from D. J. Best, 1974, Table 1, Appendix, *Division of Mathematical Statistics Technical Paper 39*, Commonwealth Scientific and Industrial Research Organization of Australia (Canberra). Reprinted with the permission of John Wiley & Sons, Inc., the Commonwealth Scientific and Industrial Research Organization of Australia, W. J. Conover, and D. J. Best.

‡The entries in this table are selected quantiles w_p of the Kendall test statistic T^* , defined by Equation 9.32 for selected values of p . Only upper quantiles are given here, but lower quantiles may be obtained from the relationship.

$$w_p = -w_{1-p}$$

Critical regions correspond to values of T^* greater than (or less than) but not including the appropriate quantile. Note that the median of T^* is 0.

§For n greater than 60, approximate quantiles of T^* may be obtained from

$$w_p \approx x_p \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

where x_p is the p^{th} quantile from the standard normal distribution given by Table A3.

TABLE A17

Quantiles of the Kolmogorov Test Statistic†‡

| | Two-sided test | | | | | | Two-sided test | | | | |
|--------------------------|----------------|------|------|------|------|-------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | p=.80 | .90 | .95 | .98 | .99 | | p=.80 | .90 | .95 | .98 | .99 |
| n=1 | .900 | .950 | .975 | .990 | .995 | n= 21 | .226 | .259 | .287 | .321 | .344 |
| 2 | .684 | .776 | .842 | .900 | .929 | 22 | .221 | .253 | .281 | .314 | .337 |
| 3 | .565 | .636 | .708 | .785 | .829 | 23 | .216 | .247 | .275 | .307 | .330 |
| 4 | .493 | .565 | .624 | .689 | .734 | 24 | .212 | .242 | .269 | .301 | .323 |
| 5 | .447 | .509 | .563 | .627 | .669 | 25 | .208 | .238 | .264 | .295 | .317 |
| 6 | .410 | .468 | .519 | .577 | .617 | 26 | .204 | .233 | .259 | .290 | .311 |
| 7 | .381 | .436 | .483 | .538 | .576 | 27 | .200 | .229 | .254 | .284 | .305 |
| 8 | .358 | .410 | .454 | .507 | .542 | 28 | .197 | .225 | .250 | .279 | .300 |
| 9 | .339 | .387 | .430 | .480 | .513 | 29 | .193 | .221 | .246 | .275 | .295 |
| 10 | .323 | .369 | .409 | .457 | .489 | 30 | .190 | .218 | .242 | .270 | .290 |
| 11 | .308 | .352 | .391 | .437 | .468 | 31 | .187 | .214 | .238 | .266 | .285 |
| 12 | .296 | .338 | .375 | .419 | .449 | 32 | .184 | .211 | .234 | .262 | .281 |
| 13 | .285 | .325 | .361 | .404 | .432 | 33 | .182 | .208 | .231 | .258 | .277 |
| 14 | .275 | .314 | .349 | .390 | .418 | 34 | .179 | .205 | .227 | .254 | .273 |
| 15 | .266 | .304 | .338 | .377 | .404 | 35 | .177 | .202 | .224 | .251 | .269 |
| 16 | .258 | .295 | .327 | .366 | .392 | 36 | .174 | .199 | .221 | .247 | .265 |
| 17 | .250 | .286 | .318 | .355 | .381 | 37 | .172 | .196 | .218 | .244 | .262 |
| 18 | .244 | .279 | .309 | .346 | .371 | 38 | .170 | .194 | .215 | .241 | .258 |
| 19 | .237 | .271 | .301 | .337 | .361 | 39 | .168 | .191 | .213 | .238 | .255 |
| 20 | .232 | .265 | .294 | .329 | .352 | 40 | .165 | .189 | .210 | .235 | .252 |
| Approximation for n > 40 | | | | | | | $\frac{1.07}{\sqrt{n}}$ | $\frac{1.22}{\sqrt{n}}$ | $\frac{1.36}{\sqrt{n}}$ | $\frac{1.52}{\sqrt{n}}$ | $\frac{1.63}{\sqrt{n}}$ |

†Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from L. H. Miller, 1956, *Journal of the American Statistical Association*, 51: 111-121 (Appendix). Reprinted with the permission of John Wiley & Sons, Inc., the American Statistical Association, and the authors.

‡The entries in this table are selected quantiles w_p of the Kolmogorov test statistic T^* as defined by Equation 9.37 for the two-sided test. Reject H_0 at the level α if T^* exceeds the $1-\alpha$ quantile given in this table. These quantiles are exact for $n \leq 40$. A better approximation for $n > 40$ results if $(n + \sqrt{n}/10)^{1/2}$ is used instead of \sqrt{n} in the denominator.

TABLE A18
Quantiles of the Lilliefors Test Statistic for Normality†‡

| | $p=.80$ | $.85$ | $.90$ | $.95$ | $.99$ |
|---------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| Sample size $n = 4$ | .300 | .319 | .352 | .381 | .417 |
| 5 | .285 | .299 | .315 | .337 | .405 |
| 6 | .265 | .277 | .294 | .319 | .364 |
| 7 | .247 | .258 | .276 | .300 | .348 |
| 8 | .233 | .244 | .261 | .285 | .331 |
| 9 | .223 | .233 | .249 | .271 | .311 |
| 10 | .215 | .224 | .239 | .258 | .294 |
| 11 | .206 | .217 | .230 | .249 | .284 |
| 12 | .199 | .212 | .223 | .242 | .275 |
| 13 | .190 | .202 | .214 | .234 | .268 |
| 14 | .183 | .194 | .207 | .227 | .261 |
| 15 | .177 | .187 | .201 | .220 | .257 |
| 16 | .173 | .182 | .195 | .213 | .250 |
| 17 | .169 | .177 | .189 | .206 | .245 |
| 18 | .166 | .173 | .184 | .200 | .239 |
| 19 | .163 | .169 | .179 | .195 | .235 |
| 20 | .160 | .166 | .174 | .190 | .231 |
| 25 | .142 | .147 | .158 | .173 | .200 |
| 30 | .131 | .136 | .144 | .161 | .187 |
| Over 30 | $\frac{.736}{\sqrt{n}}$ | $\frac{.768}{\sqrt{n}}$ | $\frac{.805}{\sqrt{n}}$ | $\frac{.886}{\sqrt{n}}$ | $\frac{1.031}{\sqrt{n}}$ |

†Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from H. W. Lilliefors, 1967, *Journal of the American Statistical Association*, 62: 399-402. Reprinted with the permission of John Wiley & Sons, Inc., the American Statistical Association, and the authors.

‡The entries in this table are the approximate quantiles w_p of the Lilliefors test statistic T^* as defined by Equation 9.39. Reject H_0 at the level α if T^* exceeds $w_{1-\alpha}$ for the particular sample size n .

TABLE A19
Coefficients a_i Used in the Shapiro-Wilk Test for Normality*

| $i \backslash n$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0.7071 | 0.6872 | 0.6646 | 0.6431 | 0.6233 | 0.6052 | 0.5888 | 0.5739 | 0.5601 | 0.5475 | 0.5359 | 0.5251 | 0.5150 | 0.5056 | 0.4968 | 0.4886 |
| 2 | | 0.1677 | 0.2413 | 0.2806 | 0.3031 | 0.3164 | 0.3244 | 0.3291 | 0.3315 | 0.3325 | 0.3325 | 0.3318 | 0.3306 | 0.3290 | 0.3273 | 0.3253 |
| 3 | | | | 0.0875 | 0.1401 | 0.1743 | 0.1976 | 0.2141 | 0.2260 | 0.2347 | 0.2412 | 0.2460 | 0.2495 | 0.2521 | 0.2540 | 0.2553 |
| 4 | | | | | | 0.0561 | 0.0947 | 0.1224 | 0.1429 | 0.1586 | 0.1707 | 0.1802 | 0.1878 | 0.1939 | 0.1988 | 0.2027 |
| 5 | | | | | | | | 0.0399 | 0.0695 | 0.0922 | 0.1099 | 0.1240 | 0.1353 | 0.1447 | 0.1524 | 0.1587 |
| 6 | | | | | | | | | | 0.0303 | 0.0539 | 0.0727 | 0.0880 | 0.1005 | 0.1109 | 0.1197 |
| 7 | | | | | | | | | | | | 0.0240 | 0.0433 | 0.0593 | 0.0725 | 0.0837 |
| 8 | | | | | | | | | | | | | | 0.0196 | 0.0359 | 0.0496 |
| 9 | | | | | | | | | | | | | | | | 0.0163 |

(Table A19 continued on next page.)

Table A19 (Cont'd)

| i | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0.4096 | 0.4068 | 0.4040 | 0.4015 | 0.3989 | 0.3964 | 0.3940 | 0.3917 | 0.3894 | 0.3872 | 0.3850 | 0.3830 | 0.3808 | 0.3789 | 0.3770 | 0.3751 |
| 2 | 0.2834 | 0.2813 | 0.2794 | 0.2774 | 0.2755 | 0.2737 | 0.2719 | 0.2701 | 0.2684 | 0.2667 | 0.2651 | 0.2635 | 0.2620 | 0.2604 | 0.2589 | 0.2574 |
| 3 | 0.2427 | 0.2415 | 0.2403 | 0.2391 | 0.2380 | 0.2368 | 0.2357 | 0.2345 | 0.2334 | 0.2323 | 0.2313 | 0.2302 | 0.2291 | 0.2281 | 0.2271 | 0.2260 |
| 4 | 0.2127 | 0.2121 | 0.2116 | 0.2110 | 0.2104 | 0.2098 | 0.2091 | 0.2085 | 0.2078 | 0.2072 | 0.2065 | 0.2058 | 0.2052 | 0.2045 | 0.2038 | 0.2032 |
| 5 | 0.1883 | 0.1883 | 0.1883 | 0.1881 | 0.1880 | 0.1878 | 0.1876 | 0.1874 | 0.1871 | 0.1868 | 0.1865 | 0.1862 | 0.1859 | 0.1855 | 0.1851 | 0.1847 |
| 6 | 0.1673 | 0.1678 | 0.1683 | 0.1686 | 0.1689 | 0.1691 | 0.1693 | 0.1694 | 0.1695 | 0.1695 | 0.1695 | 0.1695 | 0.1695 | 0.1693 | 0.1692 | 0.1691 |
| 7 | 0.1487 | 0.1496 | 0.1505 | 0.1513 | 0.1520 | 0.1526 | 0.1531 | 0.1535 | 0.1539 | 0.1542 | 0.1545 | 0.1548 | 0.1550 | 0.1551 | 0.1553 | 0.1554 |
| 8 | 0.1317 | 0.1331 | 0.1344 | 0.1356 | 0.1366 | 0.1376 | 0.1384 | 0.1392 | 0.1398 | 0.1405 | 0.1410 | 0.1415 | 0.1420 | 0.1423 | 0.1427 | 0.1430 |
| 9 | 0.1160 | 0.1179 | 0.1196 | 0.1211 | 0.1225 | 0.1237 | 0.1249 | 0.1259 | 0.1269 | 0.1278 | 0.1286 | 0.1293 | 0.1300 | 0.1306 | 0.1312 | 0.1317 |
| 10 | 0.1013 | 0.1036 | 0.1056 | 0.1075 | 0.1092 | 0.1108 | 0.1123 | 0.1136 | 0.1149 | 0.1160 | 0.1170 | 0.1180 | 0.1189 | 0.1197 | 0.1205 | 0.1212 |
| 11 | 0.0873 | 0.0900 | 0.0924 | 0.0947 | 0.0967 | 0.0986 | 0.1004 | 0.1020 | 0.1035 | 0.1049 | 0.1062 | 0.1073 | 0.1085 | 0.1095 | 0.1105 | 0.1113 |
| 12 | 0.0739 | 0.0770 | 0.0798 | 0.0824 | 0.0848 | 0.0870 | 0.0891 | 0.0909 | 0.0927 | 0.0943 | 0.0959 | 0.0972 | 0.0986 | 0.0998 | 0.1010 | 0.1020 |
| 13 | 0.0610 | 0.0645 | 0.0677 | 0.0706 | 0.0733 | 0.0759 | 0.0782 | 0.0804 | 0.0824 | 0.0842 | 0.0860 | 0.0876 | 0.0892 | 0.0906 | 0.0919 | 0.0932 |
| 14 | 0.0484 | 0.0523 | 0.0559 | 0.0592 | 0.0622 | 0.0651 | 0.0677 | 0.0701 | 0.0724 | 0.0745 | 0.0765 | 0.0783 | 0.0801 | 0.0817 | 0.0832 | 0.0846 |
| 15 | 0.0361 | 0.0404 | 0.0444 | 0.0481 | 0.0515 | 0.0546 | 0.0575 | 0.0602 | 0.0628 | 0.0651 | 0.0673 | 0.0694 | 0.0713 | 0.0731 | 0.0748 | 0.0764 |
| 16 | 0.0239 | 0.0287 | 0.0331 | 0.0372 | 0.0409 | 0.0444 | 0.0476 | 0.0506 | 0.0534 | 0.0560 | 0.0584 | 0.0607 | 0.0628 | 0.0648 | 0.0667 | 0.0685 |
| 17 | 0.0119 | 0.0172 | 0.0220 | 0.0264 | 0.0305 | 0.0343 | 0.0379 | 0.0411 | 0.0442 | 0.0471 | 0.0497 | 0.0522 | 0.0546 | 0.0568 | 0.0588 | 0.0608 |
| 18 | | 0.0057 | 0.0110 | 0.0158 | 0.0203 | 0.0244 | 0.0283 | 0.0318 | 0.0352 | 0.0383 | 0.0412 | 0.0439 | 0.0465 | 0.0489 | 0.0511 | 0.0532 |
| 19 | | | | 0.0053 | 0.0101 | 0.0146 | 0.0188 | 0.0227 | 0.0263 | 0.0296 | 0.0328 | 0.0357 | 0.0385 | 0.0411 | 0.0436 | 0.0459 |
| 20 | | | | | | 0.0049 | 0.0094 | 0.0136 | 0.0175 | 0.0211 | 0.0245 | 0.0277 | 0.0307 | 0.0335 | 0.0361 | 0.0386 |
| 21 | | | | | | | | 0.0045 | 0.0087 | 0.0126 | 0.0163 | 0.0197 | 0.0229 | 0.0259 | 0.0288 | 0.0314 |
| 22 | | | | | | | | | | 0.0042 | 0.0081 | 0.0118 | 0.0153 | 0.0185 | 0.0215 | 0.0244 |
| 23 | | | | | | | | | | | | 0.0039 | 0.0076 | 0.0111 | 0.0143 | 0.0174 |
| 24 | | | | | | | | | | | | | | 0.0037 | 0.0071 | 0.0104 |
| 25 | | | | | | | | | | | | | | | | 0.0035 |

*Hahn, G. J., and S. S. Shapiro, 1967, *Statistical Models in Engineering*, (Table IX, pp. 330-331), ©John Wiley & Sons, Inc. Reprinted with the permission of John Wiley & Sons, Inc., and the principal author.

TABLE A20

Quantiles of the Shapiro-Wilk Test Statistic†‡

| n | 0.01 | 0.02 | 0.05 | 0.10 | 0.50 | 0.90 | 0.95 | 0.98 | 0.99 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 3 | 0.753 | 0.756 | 0.767 | 0.789 | 0.959 | 0.998 | 0.999 | 1.000 | 1.000 |
| 4 | 0.687 | 0.707 | 0.748 | 0.792 | 0.935 | 0.987 | 0.992 | 0.996 | 0.997 |
| 5 | 0.686 | 0.715 | 0.762 | 0.806 | 0.927 | 0.979 | 0.986 | 0.991 | 0.993 |
| 6 | 0.713 | 0.743 | 0.788 | 0.826 | 0.927 | 0.974 | 0.981 | 0.986 | 0.989 |
| 7 | 0.730 | 0.760 | 0.803 | 0.838 | 0.928 | 0.972 | 0.979 | 0.985 | 0.988 |
| 8 | 0.749 | 0.778 | 0.818 | 0.851 | 0.932 | 0.972 | 0.978 | 0.984 | 0.987 |
| 9 | 0.764 | 0.791 | 0.829 | 0.859 | 0.935 | 0.972 | 0.978 | 0.984 | 0.986 |
| 10 | 0.781 | 0.806 | 0.842 | 0.869 | 0.938 | 0.972 | 0.978 | 0.983 | 0.986 |
| 11 | 0.792 | 0.817 | 0.850 | 0.876 | 0.940 | 0.973 | 0.979 | 0.984 | 0.986 |
| 12 | 0.805 | 0.828 | 0.859 | 0.883 | 0.943 | 0.973 | 0.979 | 0.984 | 0.986 |
| 13 | 0.814 | 0.837 | 0.866 | 0.889 | 0.945 | 0.974 | 0.979 | 0.984 | 0.986 |
| 14 | 0.825 | 0.846 | 0.874 | 0.895 | 0.947 | 0.975 | 0.980 | 0.984 | 0.986 |
| 15 | 0.835 | 0.855 | 0.881 | 0.901 | 0.950 | 0.975 | 0.980 | 0.984 | 0.987 |
| 16 | 0.844 | 0.863 | 0.887 | 0.906 | 0.952 | 0.976 | 0.981 | 0.985 | 0.987 |
| 17 | 0.851 | 0.869 | 0.892 | 0.910 | 0.954 | 0.977 | 0.981 | 0.985 | 0.987 |
| 18 | 0.858 | 0.874 | 0.897 | 0.914 | 0.956 | 0.978 | 0.982 | 0.986 | 0.988 |
| 19 | 0.863 | 0.879 | 0.901 | 0.917 | 0.957 | 0.978 | 0.982 | 0.986 | 0.988 |
| 20 | 0.868 | 0.884 | 0.905 | 0.920 | 0.959 | 0.979 | 0.983 | 0.986 | 0.988 |
| 21 | 0.873 | 0.888 | 0.908 | 0.923 | 0.960 | 0.980 | 0.983 | 0.987 | 0.989 |
| 22 | 0.878 | 0.892 | 0.911 | 0.926 | 0.961 | 0.980 | 0.984 | 0.987 | 0.989 |
| 23 | 0.881 | 0.895 | 0.914 | 0.928 | 0.962 | 0.981 | 0.984 | 0.987 | 0.989 |
| 24 | 0.884 | 0.898 | 0.916 | 0.930 | 0.963 | 0.981 | 0.984 | 0.987 | 0.989 |
| 25 | 0.888 | 0.901 | 0.918 | 0.931 | 0.964 | 0.981 | 0.985 | 0.988 | 0.989 |
| 26 | 0.891 | 0.904 | 0.920 | 0.933 | 0.965 | 0.982 | 0.985 | 0.988 | 0.989 |
| 27 | 0.894 | 0.906 | 0.923 | 0.935 | 0.965 | 0.982 | 0.985 | 0.988 | 0.990 |
| 28 | 0.896 | 0.908 | 0.924 | 0.936 | 0.966 | 0.982 | 0.985 | 0.988 | 0.990 |
| 29 | 0.898 | 0.910 | 0.926 | 0.937 | 0.966 | 0.982 | 0.985 | 0.988 | 0.990 |
| 30 | 0.900 | 0.912 | 0.927 | 0.939 | 0.967 | 0.983 | 0.985 | 0.988 | 0.990 |
| 31 | 0.902 | 0.914 | 0.929 | 0.940 | 0.967 | 0.983 | 0.986 | 0.988 | 0.990 |
| 32 | 0.904 | 0.915 | 0.930 | 0.941 | 0.968 | 0.983 | 0.986 | 0.988 | 0.990 |
| 33 | 0.906 | 0.917 | 0.931 | 0.942 | 0.968 | 0.983 | 0.986 | 0.989 | 0.990 |
| 34 | 0.908 | 0.919 | 0.933 | 0.943 | 0.969 | 0.983 | 0.986 | 0.989 | 0.990 |
| 35 | 0.910 | 0.920 | 0.934 | 0.944 | 0.969 | 0.984 | 0.986 | 0.989 | 0.990 |
| 36 | 0.912 | 0.922 | 0.935 | 0.945 | 0.970 | 0.984 | 0.986 | 0.989 | 0.990 |
| 37 | 0.914 | 0.924 | 0.936 | 0.946 | 0.970 | 0.984 | 0.987 | 0.989 | 0.990 |
| 38 | 0.916 | 0.925 | 0.938 | 0.947 | 0.971 | 0.984 | 0.987 | 0.989 | 0.990 |
| 39 | 0.917 | 0.927 | 0.939 | 0.948 | 0.971 | 0.984 | 0.987 | 0.989 | 0.991 |
| 40 | 0.919 | 0.928 | 0.940 | 0.949 | 0.972 | 0.985 | 0.987 | 0.989 | 0.991 |
| 41 | 0.920 | 0.929 | 0.941 | 0.950 | 0.972 | 0.985 | 0.987 | 0.989 | 0.991 |
| 42 | 0.922 | 0.930 | 0.942 | 0.951 | 0.972 | 0.985 | 0.987 | 0.989 | 0.991 |
| 43 | 0.923 | 0.932 | 0.943 | 0.951 | 0.973 | 0.985 | 0.987 | 0.990 | 0.991 |
| 44 | 0.924 | 0.933 | 0.944 | 0.952 | 0.973 | 0.985 | 0.987 | 0.990 | 0.991 |
| 45 | 0.926 | 0.934 | 0.945 | 0.953 | 0.973 | 0.985 | 0.988 | 0.990 | 0.991 |
| 46 | 0.927 | 0.935 | 0.945 | 0.953 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |
| 47 | 0.928 | 0.936 | 0.946 | 0.954 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |
| 48 | 0.929 | 0.937 | 0.947 | 0.954 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |
| 49 | 0.929 | 0.937 | 0.947 | 0.955 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |
| 50 | 0.930 | 0.938 | 0.947 | 0.955 | 0.974 | 0.985 | 0.988 | 0.990 | 0.991 |

(Footnote to Table A20 at bottom of next page).—→

TABLE A21

**Table of Critical Values for T_n (One-Sided Test)
When Standard Deviation Is Calculated
from the Same Sample***

| Number of observations
n | 5%
Significance level | 2.5%
Significance level | 1%
Significance level |
|-------------------------------|--------------------------|----------------------------|--------------------------|
| 3 | 1.15 | 1.15 | 1.15 |
| 4 | 1.46 | 1.48 | 1.49 |
| 5 | 1.67 | 1.71 | 1.75 |
| 6 | 1.82 | 1.89 | 1.94 |
| 7 | 1.94 | 2.02 | 2.10 |
| 8 | 2.03 | 2.13 | 2.22 |
| 9 | 2.11 | 2.21 | 2.32 |
| 10 | 2.18 | 2.29 | 2.41 |
| 11 | 2.23 | 2.36 | 2.48 |
| 12 | 2.29 | 2.41 | 2.55 |
| 13 | 2.33 | 2.46 | 2.61 |
| 14 | 2.37 | 2.51 | 2.66 |
| 15 | 2.41 | 2.55 | 2.71 |
| 16 | 2.44 | 2.59 | 2.75 |
| 17 | 2.47 | 2.62 | 2.79 |
| 18 | 2.50 | 2.65 | 2.82 |
| 19 | 2.53 | 2.68 | 2.85 |
| 20 | 2.56 | 2.71 | 2.88 |
| 21 | 2.58 | 2.73 | 2.91 |
| 22 | 2.60 | 2.76 | 2.94 |
| 23 | 2.62 | 2.78 | 2.96 |
| 24 | 2.64 | 2.80 | 2.99 |
| 25 | 2.66 | 2.82 | 3.01 |
| 30 | 2.75 | 2.91 | |
| 35 | 2.82 | 2.98 | |
| 40 | 2.87 | 3.04 | |
| 45 | 2.92 | 3.09 | |
| 50 | 2.96 | 3.13 | |
| 60 | 3.03 | 3.20 | |
| 70 | 3.09 | 3.26 | |
| 80 | 3.14 | 3.31 | |
| 90 | 3.18 | 3.35 | |
| 100 | 3.21 | 3.38 | |

*Grubbs, F. E., 1969, *Technometrics*, 11: 1-21.
Reprinted with the permission of the American Statistical Association and the author.

(Footnote to Table 20.)

†Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from S. S. Shapiro and M. B. Wilk, 1965, *Biometrika*, 52: 591. Reprinted with the permission of John Wiley & Sons, Inc., the Biometrika Trustees, and the authors W. J. Conover and S. S. Shapiro.

‡The entries in this table are quantiles w_p of the Shapiro-Wilk Test statistic W^* given by Equation 9.40.

TABLE A22
Critical Values for w/s (Ratio of Range to Sample Standard Deviation)*

| Number of
observations
<i>n</i> | 5%
Significance
level | 1%
Significance
level | 0.5%
Significance
level |
|---------------------------------------|-----------------------------|-----------------------------|-------------------------------|
| 3 | 2.00 | 2.00 | 2.00 |
| 4 | 2.43 | 2.44 | 2.45 |
| 5 | 2.75 | 2.80 | 2.81 |
| 6 | 3.01 | 3.10 | 3.12 |
| 7 | 3.22 | 3.34 | 3.37 |
| 8 | 3.40 | 3.54 | 3.58 |
| 9 | 3.55 | 3.72 | 3.77 |
| 10 | 3.68 | 3.88 | 3.94 |
| 11 | 3.80 | 4.01 | 4.08 |
| 12 | 3.91 | 4.13 | 4.21 |
| 13 | 4.00 | 4.24 | 4.32 |
| 14 | 4.09 | 4.34 | 4.43 |
| 15 | 4.17 | 4.43 | 4.53 |
| 16 | 4.24 | 4.51 | 4.62 |
| 17 | 4.31 | 4.59 | 4.69 |
| 18 | 4.38 | 4.66 | 4.77 |
| 19 | 4.43 | 4.73 | 4.84 |
| 20 | 4.49 | 4.79 | 4.91 |
| 30 | 4.89 | 5.25 | 5.39 |
| 40 | 5.15 | 5.54 | 5.69 |
| 50 | 5.35 | 5.77 | 5.91 |
| 60 | 5.50 | 5.93 | 6.09 |
| 80 | 5.73 | 6.18 | 6.35 |
| 100 | 5.90 | 6.36 | 6.54 |
| 150 | 6.18 | 6.64 | 6.84 |
| 200 | 6.38 | 6.85 | 7.03 |
| 500 | 6.94 | 7.42 | 7.60 |
| 1000 | 7.33 | 7.80 | 7.99 |

*Selected from H. A. David, H. O. Hartley, and E. S. Pearson, 1954. "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation," *Biometrika*, 41: 482-493. Reprinted with the permission of the Biometrika Trustees and the principal author.

TABLE A23
Critical Values for Simultaneously Testing the Two
Largest or Two Smallest Observations*

| Number of
observations
<i>n</i> | 10%
Significance
level | 5%
Significance
level | 1%
Significance
level |
|---------------------------------------|------------------------------|-----------------------------|-----------------------------|
| 4 | .0031 | .0008 | .0000 |
| 5 | .0376 | .0183 | .0035 |
| 6 | .0921 | .0565 | .0186 |
| 7 | .1479 | .1020 | .0440 |
| 8 | .1994 | .1478 | .0750 |
| 9 | .2454 | .1909 | .1082 |
| 10 | .2853 | .2305 | .1415 |
| 11 | .3226 | .2666 | .1736 |
| 12 | .3552 | .2996 | .2044 |
| 13 | .3843 | .3295 | .2333 |
| 14 | .4106 | .3568 | .2605 |
| 15 | .4345 | .3818 | .2859 |
| 16 | .4562 | .4048 | .3098 |
| 17 | .4761 | .4259 | .3321 |
| 18 | .4944 | .4455 | .3530 |
| 19 | .5113 | .4636 | .3725 |
| 20 | .5269 | .4804 | .3909 |

*Grubbs, F. E., 1969, *Technometrics*, 11: 1-21. Reprinted with the permission of the publisher and the author.

TABLE A24
Critical Values for Testing $\sqrt{b_1}$ *

| | | <i>n</i> | | | | | | | | | |
|-----------|------|----------|------|------|------|-----|-----|-----|-----|-----|--|
| Sig Level | 5† | 10† | 15† | 20† | 25 | 30 | 35 | 40 | 50 | 60 | |
| 1% | 1.34 | 1.31 | 1.20 | 1.11 | 1.06 | .98 | .92 | .87 | .79 | .72 | |
| 5% | 1.05 | .92 | .84 | .79 | .71 | .66 | .62 | .59 | .53 | .49 | |

*Grubbs, F. E., 1969, *Technometrics*, 11: 1-21. Reprinted with the permission of the American Statistical Association and the author.

†Ferguson, T. S., 1961, "On the Rejection of Outliers" in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 253-287. Reprinted with the permission of the publisher and the author.

TABLE A25
Critical Values for Testing b_2 *

| | | <i>n</i> | | | | | | | |
|-----------|------|----------|------|------|------|------|------|------|--|
| Sig Level | 5† | 10† | 15† | 20† | 25† | 50 | 75 | 100 | |
| 1% | 3.11 | 4.83 | 5.08 | 5.23 | 5.00 | 4.88 | 4.59 | 4.39 | |
| 5% | 2.89 | 3.85 | 4.07 | 4.15 | 4.00 | 3.99 | 3.87 | 3.77 | |

*Grubbs, F. E., 1969, *Technometrics*, 11: 1-21. Reprinted with the permission of the American Statistical Association and the author.

†Ferguson, T. S., 1961, "On the Rejection of Outliers" in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 253-287. Reprinted with the permission of the publisher and the author.

TABLE A26
 Critical Values for T'_n When Standard Deviation s_x Is
 Independent of Present Sample*

| n | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 |
|------------|-----------|------|------|------|------|------|------|------|------|
| $\nu = df$ | 1% points | | | | | | | | |
| 10 | 2.78 | 3.10 | 3.32 | 3.48 | 3.62 | 3.73 | 3.82 | 3.90 | 4.04 |
| 11 | 2.72 | 3.02 | 3.24 | 3.39 | 3.52 | 3.63 | 3.72 | 3.79 | 3.93 |
| 12 | 2.67 | 2.96 | 3.17 | 3.32 | 3.45 | 3.55 | 3.64 | 3.71 | 3.84 |
| 13 | 2.63 | 2.92 | 3.12 | 3.27 | 3.38 | 3.48 | 3.57 | 3.64 | 3.76 |
| 14 | 2.60 | 2.88 | 3.07 | 3.22 | 3.33 | 3.43 | 3.51 | 3.58 | 3.70 |
| 15 | 2.57 | 2.84 | 3.03 | 3.17 | 3.29 | 3.38 | 3.46 | 3.53 | 3.65 |
| 16 | 2.54 | 2.81 | 3.00 | 3.14 | 3.25 | 3.34 | 3.42 | 3.49 | 3.60 |
| 17 | 2.52 | 2.79 | 2.97 | 3.11 | 3.22 | 3.31 | 3.38 | 3.45 | 3.56 |
| 18 | 2.50 | 2.77 | 2.95 | 3.08 | 3.19 | 3.28 | 3.35 | 3.42 | 3.53 |
| 19 | 2.49 | 2.75 | 2.93 | 3.06 | 3.16 | 3.25 | 3.33 | 3.39 | 3.50 |
| 20 | 2.47 | 2.73 | 2.91 | 3.04 | 3.14 | 3.23 | 3.30 | 3.37 | 3.47 |
| 24 | 2.42 | 2.68 | 2.84 | 2.97 | 3.07 | 3.16 | 3.23 | 3.29 | 3.38 |
| 30 | 2.38 | 2.62 | 2.79 | 2.91 | 3.01 | 3.08 | 3.15 | 3.21 | 3.30 |
| 40 | 2.34 | 2.57 | 2.73 | 2.85 | 2.94 | 3.02 | 3.08 | 3.13 | 3.22 |
| 60 | 2.29 | 2.52 | 2.68 | 2.79 | 2.88 | 2.95 | 3.01 | 3.06 | 3.15 |
| 120 | 2.25 | 2.48 | 2.62 | 2.73 | 2.82 | 2.89 | 2.95 | 3.00 | 3.08 |
| ∞ | 2.22 | 2.43 | 2.57 | 2.68 | 2.76 | 2.83 | 2.88 | 2.93 | 3.01 |
| | 5% points | | | | | | | | |
| 10 | 2.01 | 2.27 | 2.46 | 2.60 | 2.72 | 2.81 | 2.89 | 2.96 | 3.08 |
| 11 | 1.98 | 2.24 | 2.42 | 2.56 | 2.67 | 2.76 | 2.84 | 2.91 | 3.03 |
| 12 | 1.96 | 2.21 | 2.39 | 2.52 | 2.63 | 2.72 | 2.80 | 2.87 | 2.98 |
| 13 | 1.94 | 2.19 | 2.36 | 2.50 | 2.60 | 2.69 | 2.76 | 2.83 | 2.94 |
| 14 | 1.93 | 2.17 | 2.34 | 2.47 | 2.57 | 2.66 | 2.74 | 2.80 | 2.91 |
| 15 | 1.91 | 2.15 | 2.32 | 2.45 | 2.55 | 2.64 | 2.71 | 2.77 | 2.88 |
| 16 | 1.90 | 2.14 | 2.31 | 2.43 | 2.53 | 2.62 | 2.69 | 2.75 | 2.86 |
| 17 | 1.89 | 2.13 | 2.29 | 2.42 | 2.52 | 2.60 | 2.67 | 2.73 | 2.84 |
| 18 | 1.88 | 2.11 | 2.28 | 2.40 | 2.50 | 2.58 | 2.65 | 2.71 | 2.82 |
| 19 | 1.87 | 2.11 | 2.27 | 2.39 | 2.49 | 2.57 | 2.64 | 2.70 | 2.80 |
| 20 | 1.87 | 2.10 | 2.26 | 2.38 | 2.47 | 2.56 | 2.63 | 2.68 | 2.78 |
| 24 | 1.84 | 2.07 | 2.23 | 2.34 | 2.44 | 2.52 | 2.58 | 2.64 | 2.74 |
| 30 | 1.82 | 2.04 | 2.20 | 2.31 | 2.40 | 2.48 | 2.54 | 2.60 | 2.69 |
| 40 | 1.80 | 2.02 | 2.17 | 2.28 | 2.37 | 2.44 | 2.50 | 2.56 | 2.65 |
| 60 | 1.78 | 1.99 | 2.14 | 2.25 | 2.33 | 2.41 | 2.47 | 2.52 | 2.61 |
| 120 | 1.76 | 1.96 | 2.11 | 2.22 | 2.30 | 2.37 | 2.43 | 2.48 | 2.57 |
| ∞ | 1.74 | 1.94 | 2.08 | 2.18 | 2.27 | 2.33 | 2.39 | 2.44 | 2.52 |

*The percentage points are from David, H. A., 1956, "Revised Upper Percentage Points of the Extreme Studentized Deviate from the Sample Mean," *Biometrika*, 43: 449-451. Reprinted with the permission of the Biometrika Trustees and the author.

TABLE A27
Critical Values of T'_1 and T'_n When the
Population Standard Deviation σ Is Known*

| Number of
observations
n | 5%
Significance
level | 1%
Significance
level | 0.5%
Significance
level |
|----------------------------------|-----------------------------|-----------------------------|-------------------------------|
| 2 | 1.39 | 1.82 | 1.99 |
| 3 | 1.74 | 2.22 | 2.40 |
| 4 | 1.94 | 2.43 | 2.62 |
| 5 | 2.08 | 2.57 | 2.76 |
| 6 | 2.18 | 2.68 | 2.87 |
| 7 | 2.27 | 2.76 | 2.95 |
| 8 | 2.33 | 2.83 | 3.02 |
| 9 | 2.39 | 2.88 | 3.07 |
| 10 | 2.44 | 2.93 | 3.12 |
| 11 | 2.48 | 2.97 | 3.16 |
| 12 | 2.52 | 3.01 | 3.20 |
| 13 | 2.56 | 3.04 | 3.23 |
| 14 | 2.59 | 3.07 | 3.26 |
| 15 | 2.62 | 3.10 | 3.29 |
| 16 | 2.64 | 3.12 | 3.31 |
| 17 | 2.67 | 3.15 | 3.33 |
| 18 | 2.69 | 3.17 | 3.36 |
| 19 | 2.71 | 3.19 | 3.38 |
| 20 | 2.73 | 3.21 | 3.39 |
| 21 | 2.75 | 3.22 | 3.41 |
| 22 | 2.77 | 3.24 | 3.42 |
| 23 | 2.78 | 3.26 | 3.44 |
| 24 | 2.80 | 3.27 | 3.45 |
| 25 | 2.81 | 3.28 | 3.46 |

*Grubbs, F. E., 1969, *Technometrics*, 11: 1-21.
 Reprinted with the permission of the publisher and the
 author.

TABLE A28
Critical Values for Screening Studentized Residuals*

$\alpha = 0.01$

P, the number of parameters

| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|------|
| 6 | 7.45 | | | | | | | | | | | | | | | | | | | | |
| 7 | 6.32 | 7.75 | | | | | | | | | | | | | | | | | | | |
| 8 | 5.69 | 6.51 | 8.02 | | | | | | | | | | | | | | | | | | |
| 9 | 5.30 | 5.82 | 6.68 | 8.27 | | | | | | | | | | | | | | | | | |
| 10 | 5.04 | 5.40 | 5.94 | 6.83 | 8.49 | | | | | | | | | | | | | | | | |
| 11 | 4.84 | 5.11 | 5.49 | 6.05 | 6.97 | 8.70 | | | | | | | | | | | | | | | |
| 12 | 4.70 | 4.91 | 5.18 | 5.57 | 6.15 | 7.10 | 8.89 | | | | | | | | | | | | | | |
| 13 | 4.59 | 4.76 | 4.97 | 5.25 | 5.65 | 6.24 | 7.22 | 9.07 | | | | | | | | | | | | | |
| 14 | 4.51 | 4.64 | 4.80 | 5.02 | 5.31 | 5.72 | 6.33 | 7.34 | 9.24 | | | | | | | | | | | | |
| 15 | 4.44 | 4.55 | 4.68 | 4.85 | 5.07 | 5.37 | 5.78 | 6.41 | 7.44 | 9.40 | | | | | | | | | | | |
| 16 | 4.38 | 4.47 | 4.58 | 4.72 | 4.89 | 5.12 | 5.42 | 5.85 | 6.49 | 7.55 | 9.55 | | | | | | | | | | |
| 17 | 4.33 | 4.41 | 4.51 | 4.62 | 4.76 | 4.93 | 5.16 | 5.47 | 5.91 | 6.56 | 7.64 | 9.70 | | | | | | | | | |
| 18 | 4.29 | 4.36 | 4.44 | 4.54 | 4.65 | 4.79 | 4.97 | 5.21 | 5.52 | 5.96 | 6.63 | 7.74 | 9.83 | | | | | | | | |
| 19 | 4.26 | 4.32 | 4.39 | 4.47 | 4.57 | 4.68 | 4.83 | 5.01 | 5.25 | 5.56 | 6.02 | 6.70 | 7.82 | 9.97 | | | | | | | |
| 20 | 4.23 | 4.28 | 4.34 | 4.41 | 4.50 | 4.59 | 4.71 | 4.86 | 5.04 | 5.28 | 5.61 | 6.07 | 6.76 | 7.91 | 10.09 | | | | | | |
| 21 | 4.21 | 4.25 | 4.31 | 4.37 | 4.44 | 4.52 | 4.62 | 4.74 | 4.89 | 5.08 | 5.32 | 5.65 | 6.12 | 6.82 | 7.99 | 10.22 | | | | | |
| 22 | 4.18 | 4.23 | 4.27 | 4.33 | 4.39 | 4.46 | 4.55 | 4.65 | 4.77 | 4.92 | 5.11 | 5.36 | 5.69 | 6.16 | 6.88 | 8.07 | 10.33 | | | | |
| 23 | 4.17 | 4.20 | 4.25 | 4.29 | 4.35 | 4.41 | 4.48 | 4.57 | 4.67 | 4.80 | 4.95 | 5.14 | 5.39 | 5.73 | 6.21 | 6.93 | 8.14 | 10.45 | | | |
| 24 | 4.15 | 4.18 | 4.22 | 4.26 | 4.31 | 4.37 | 4.43 | 4.51 | 4.59 | 4.70 | 4.82 | 4.97 | 5.17 | 5.42 | 5.77 | 6.25 | 6.99 | 8.21 | 10.56 | | |
| 25 | 4.14 | 4.17 | 4.20 | 4.24 | 4.28 | 4.33 | 4.39 | 4.45 | 4.53 | 4.61 | 4.72 | 4.84 | 5.00 | 5.20 | 5.45 | 5.80 | 6.29 | 7.04 | 8.28 | 10.66 | |
| 26 | 4.12 | 4.15 | 4.18 | 4.22 | 4.26 | 4.30 | 4.35 | 4.41 | 4.47 | 4.55 | 4.63 | 4.74 | 4.87 | 5.02 | 5.22 | 5.48 | 5.84 | 6.33 | 7.09 | 8.35 | |
| 27 | 4.11 | 4.14 | 4.17 | 4.20 | 4.23 | 4.27 | 4.32 | 4.37 | 4.42 | 4.49 | 4.56 | 4.65 | 4.76 | 4.89 | 5.05 | 5.25 | 5.51 | 5.87 | 6.37 | 7.14 | |
| 28 | 4.10 | 4.13 | 4.15 | 4.18 | 4.21 | 4.25 | 4.29 | 4.33 | 4.38 | 4.44 | 4.51 | 4.58 | 4.67 | 4.78 | 4.91 | 5.07 | 5.28 | 5.54 | 5.90 | 6.41 | |
| 29 | 4.09 | 4.11 | 4.14 | 4.17 | 4.19 | 4.23 | 4.26 | 4.30 | 4.35 | 4.40 | 4.46 | 4.52 | 4.60 | 4.69 | 4.80 | 4.93 | 5.09 | 5.30 | 5.57 | 5.93 | |
| 30 | 4.08 | 4.11 | 4.13 | 4.15 | 4.18 | 4.21 | 4.24 | 4.28 | 4.32 | 4.36 | 4.41 | 4.47 | 4.54 | 4.62 | 4.71 | 4.82 | 4.95 | 5.12 | 5.32 | 5.59 | |
| 31 | 4.08 | 4.10 | 4.12 | 4.14 | 4.16 | 4.19 | 4.22 | 4.25 | 4.29 | 4.33 | 4.38 | 4.43 | 4.49 | 4.56 | 4.63 | 4.73 | 4.84 | 4.97 | 5.14 | 5.35 | |
| 32 | 4.07 | 4.09 | 4.11 | 4.13 | 4.15 | 4.18 | 4.20 | 4.23 | 4.27 | 4.30 | 4.35 | 4.39 | 4.44 | 4.50 | 4.57 | 4.65 | 4.74 | 4.86 | 4.99 | 5.16 | |
| 33 | 4.07 | 4.08 | 4.10 | 4.12 | 4.14 | 4.16 | 4.19 | 4.22 | 4.25 | 4.28 | 4.32 | 4.36 | 4.40 | 4.46 | 4.52 | 4.59 | 4.67 | 4.76 | 4.87 | 5.01 | |
| 34 | 4.06 | 4.08 | 4.09 | 4.11 | 4.13 | 4.15 | 4.18 | 4.20 | 4.23 | 4.26 | 4.29 | 4.33 | 4.37 | 4.42 | 4.47 | 4.53 | 4.60 | 4.68 | 4.78 | 4.89 | |
| 35 | 4.06 | 4.07 | 4.09 | 4.10 | 4.12 | 4.14 | 4.16 | 4.19 | 4.21 | 4.24 | 4.27 | 4.30 | 4.34 | 4.38 | 4.43 | 4.48 | 4.54 | 4.61 | 4.70 | 4.79 | |
| 36 | 4.05 | 4.07 | 4.08 | 4.10 | 4.11 | 4.13 | 4.15 | 4.17 | 4.20 | 4.22 | 4.25 | 4.28 | 4.32 | 4.35 | 4.40 | 4.44 | 4.50 | 4.56 | 4.63 | 4.71 | |
| 37 | 4.05 | 4.06 | 4.08 | 4.09 | 4.11 | 4.12 | 4.14 | 4.16 | 4.18 | 4.21 | 4.23 | 4.26 | 4.29 | 4.33 | 4.36 | 4.41 | 4.46 | 4.51 | 4.57 | 4.64 | |
| 38 | 4.05 | 4.06 | 4.07 | 4.08 | 4.10 | 4.12 | 4.13 | 4.15 | 4.17 | 4.19 | 4.22 | 4.24 | 4.27 | 4.30 | 4.34 | 4.38 | 4.42 | 4.47 | 4.52 | 4.58 | |
| 39 | 4.04 | 4.05 | 4.07 | 4.08 | 4.09 | 4.11 | 4.13 | 4.14 | 4.16 | 4.18 | 4.20 | 4.23 | 4.25 | 4.28 | 4.31 | 4.35 | 4.39 | 4.43 | 4.48 | 4.53 | |
| 40 | 4.04 | 4.05 | 4.06 | 4.08 | 4.09 | 4.10 | 4.12 | 4.13 | 4.15 | 4.17 | 4.19 | 4.21 | 4.24 | 4.26 | 4.29 | 4.32 | 4.36 | 4.40 | 4.44 | 4.49 | |
| 41 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.10 | 4.11 | 4.13 | 4.14 | 4.16 | 4.18 | 4.20 | 4.22 | 4.25 | 4.27 | 4.30 | 4.33 | 4.37 | 4.41 | 4.45 | |
| 42 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.11 | 4.12 | 4.14 | 4.15 | 4.17 | 4.19 | 4.21 | 4.23 | 4.26 | 4.28 | 4.31 | 4.34 | 4.38 | 4.42 | |
| 43 | 4.03 | 4.04 | 4.05 | 4.06 | 4.08 | 4.09 | 4.10 | 4.11 | 4.13 | 4.14 | 4.16 | 4.18 | 4.20 | 4.22 | 4.24 | 4.26 | 4.29 | 4.32 | 4.35 | 4.39 | |
| 44 | 4.03 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.11 | 4.12 | 4.14 | 4.15 | 4.17 | 4.19 | 4.21 | 4.23 | 4.25 | 4.27 | 4.30 | 4.33 | 4.36 | |
| 45 | 4.03 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.10 | 4.12 | 4.13 | 4.14 | 4.16 | 4.18 | 4.19 | 4.21 | 4.23 | 4.26 | 4.28 | 4.31 | 4.34 | |
| 46 | 4.03 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.10 | 4.11 | 4.12 | 4.14 | 4.15 | 4.17 | 4.18 | 4.20 | 4.22 | 4.24 | 4.27 | 4.29 | 4.32 | |
| 47 | 4.03 | 4.04 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.11 | 4.12 | 4.13 | 4.14 | 4.16 | 4.17 | 4.19 | 4.21 | 4.23 | 4.25 | 4.27 | 4.30 | |
| 48 | 4.03 | 4.03 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.10 | 4.11 | 4.12 | 4.14 | 4.15 | 4.17 | 4.18 | 4.20 | 4.22 | 4.24 | 4.26 | 4.28 | |
| 49 | 4.03 | 4.03 | 4.04 | 4.05 | 4.06 | 4.07 | 4.08 | 4.09 | 4.10 | 4.11 | 4.12 | 4.13 | 4.14 | 4.16 | 4.17 | 4.19 | 4.21 | 4.23 | 4.24 | 4.27 | |
| 50 | 4.03 | 4.03 | 4.04 | 4.05 | 4.06 | 4.06 | 4.07 | 4.08 | 4.09 | 4.10 | 4.11 | 4.12 | 4.13 | 4.14 | 4.15 | 4.17 | 4.18 | 4.20 | 4.21 | 4.23 | 4.25 |

(Table A28 continued on next page.)

TABLE A28 (Cont'd)

| $\alpha = 0.05$ | | | | | | | | | | | | | | | | | | | | |
|-----------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| P, the number of parameters | | | | | | | | | | | | | | | | | | | | |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 6 | 4.81 | | | | | | | | | | | | | | | | | | | |
| 7 | 4.36 | 5.02 | | | | | | | | | | | | | | | | | | |
| 8 | 4.09 | 4.50 | 5.21 | | | | | | | | | | | | | | | | | |
| 9 | 3.93 | 4.20 | 4.63 | 5.38 | | | | | | | | | | | | | | | | |
| 10 | 3.82 | 4.01 | 4.29 | 4.74 | 5.54 | | | | | | | | | | | | | | | |
| 11 | 3.74 | 3.88 | 4.09 | 4.38 | 4.85 | 5.69 | | | | | | | | | | | | | | |
| 12 | 3.68 | 3.79 | 3.95 | 4.15 | 4.46 | 4.95 | 5.82 | | | | | | | | | | | | | |
| 13 | 3.63 | 3.73 | 3.84 | 4.00 | 4.22 | 4.54 | 5.04 | 5.95 | | | | | | | | | | | | |
| 14 | 3.60 | 3.68 | 3.77 | 3.89 | 4.06 | 4.28 | 4.61 | 5.13 | 6.07 | | | | | | | | | | | |
| 15 | 3.57 | 3.64 | 3.72 | 3.81 | 3.94 | 4.10 | 4.33 | 4.67 | 5.21 | 6.18 | | | | | | | | | | |
| 16 | 3.55 | 3.61 | 3.67 | 3.75 | 3.85 | 3.98 | 4.15 | 4.39 | 4.73 | 5.29 | 6.29 | | | | | | | | | |
| 17 | 3.53 | 3.58 | 3.64 | 3.70 | 3.79 | 3.89 | 4.02 | 4.20 | 4.44 | 4.79 | 5.36 | 6.39 | | | | | | | | |
| 18 | 3.52 | 3.56 | 3.61 | 3.67 | 3.74 | 3.82 | 3.92 | 4.06 | 4.24 | 4.48 | 4.85 | 5.43 | 6.49 | | | | | | | |
| 19 | 3.51 | 3.55 | 3.59 | 3.64 | 3.70 | 3.77 | 3.85 | 3.96 | 4.10 | 4.28 | 4.53 | 4.90 | 5.50 | 6.58 | | | | | | |
| 20 | 3.50 | 3.53 | 3.57 | 3.61 | 3.66 | 3.72 | 3.79 | 3.88 | 3.99 | 4.13 | 4.32 | 4.57 | 4.95 | 5.56 | 6.67 | | | | | |
| 21 | 3.49 | 3.52 | 3.55 | 3.59 | 3.64 | 3.69 | 3.75 | 3.82 | 3.91 | 4.02 | 4.16 | 4.35 | 4.61 | 5.00 | 5.62 | 6.76 | | | | |
| 22 | 3.49 | 3.51 | 3.54 | 3.58 | 3.61 | 3.66 | 3.71 | 3.77 | 3.85 | 3.94 | 4.05 | 4.19 | 4.39 | 4.65 | 5.05 | 5.68 | 6.84 | | | |
| 23 | 3.48 | 3.51 | 3.53 | 3.56 | 3.60 | 3.64 | 3.68 | 3.73 | 3.80 | 3.87 | 3.96 | 4.08 | 4.22 | 4.42 | 4.69 | 5.09 | 5.74 | 6.92 | | |
| 24 | 3.48 | 3.50 | 3.52 | 3.55 | 3.58 | 3.62 | 3.66 | 3.70 | 3.75 | 3.82 | 3.89 | 3.99 | 4.10 | 4.25 | 4.45 | 4.73 | 5.13 | 5.79 | 7.00 | |
| 25 | 3.48 | 3.50 | 3.52 | 3.54 | 3.57 | 3.60 | 3.63 | 3.67 | 3.72 | 3.78 | 3.84 | 3.92 | 4.01 | 4.13 | 4.28 | 4.48 | 4.76 | 5.18 | 5.84 | 7.07 |
| 26 | 3.47 | 3.49 | 3.51 | 3.53 | 3.56 | 3.59 | 3.62 | 3.65 | 3.69 | 3.74 | 3.80 | 3.86 | 3.94 | 4.03 | 4.15 | 4.31 | 4.51 | 4.80 | 5.22 | 5.90 |
| 27 | 3.47 | 3.49 | 3.51 | 3.53 | 3.55 | 3.57 | 3.60 | 3.63 | 3.67 | 3.71 | 3.76 | 3.81 | 3.88 | 3.96 | 4.06 | 4.18 | 4.33 | 4.54 | 4.83 | 5.26 |
| 28 | 3.47 | 3.49 | 3.50 | 3.52 | 3.54 | 3.57 | 3.59 | 3.62 | 3.65 | 3.69 | 3.73 | 3.78 | 3.83 | 3.90 | 3.98 | 4.08 | 4.20 | 4.36 | 4.57 | 4.86 |
| 29 | 3.47 | 3.48 | 3.50 | 3.52 | 3.54 | 3.56 | 3.58 | 3.61 | 3.63 | 3.67 | 3.70 | 3.75 | 3.79 | 3.85 | 3.92 | 4.00 | 4.10 | 4.22 | 4.38 | 4.59 |
| 30 | 3.47 | 3.48 | 3.50 | 3.51 | 3.53 | 3.55 | 3.57 | 3.59 | 3.62 | 3.65 | 3.68 | 3.72 | 3.76 | 3.81 | 3.87 | 3.94 | 4.02 | 4.12 | 4.24 | 4.40 |
| 31 | 3.47 | 3.48 | 3.50 | 3.51 | 3.53 | 3.54 | 3.56 | 3.58 | 3.61 | 3.63 | 3.66 | 3.70 | 3.73 | 3.78 | 3.83 | 3.88 | 3.95 | 4.04 | 4.14 | 4.26 |
| 32 | 3.47 | 3.48 | 3.49 | 3.51 | 3.52 | 3.54 | 3.56 | 3.58 | 3.60 | 3.62 | 3.65 | 3.68 | 3.71 | 3.75 | 3.79 | 3.84 | 3.90 | 3.97 | 4.05 | 4.16 |
| 33 | 3.47 | 3.48 | 3.49 | 3.51 | 3.52 | 3.54 | 3.55 | 3.57 | 3.59 | 3.61 | 3.63 | 3.66 | 3.69 | 3.72 | 3.76 | 3.81 | 3.86 | 3.92 | 3.99 | 4.07 |
| 34 | 3.47 | 3.48 | 3.49 | 3.50 | 3.52 | 3.53 | 3.55 | 3.56 | 3.58 | 3.60 | 3.62 | 3.65 | 3.67 | 3.70 | 3.74 | 3.78 | 3.82 | 3.87 | 3.93 | 4.00 |
| 35 | 3.47 | 3.48 | 3.49 | 3.50 | 3.52 | 3.53 | 3.54 | 3.56 | 3.57 | 3.59 | 3.61 | 3.63 | 3.66 | 3.69 | 3.72 | 3.75 | 3.79 | 3.83 | 3.89 | 3.95 |
| 36 | 3.47 | 3.48 | 3.49 | 3.50 | 3.51 | 3.53 | 3.54 | 3.55 | 3.57 | 3.59 | 3.60 | 3.62 | 3.65 | 3.67 | 3.70 | 3.73 | 3.76 | 3.80 | 3.85 | 3.90 |
| 37 | 3.47 | 3.48 | 3.49 | 3.50 | 3.51 | 3.52 | 3.54 | 3.55 | 3.56 | 3.58 | 3.60 | 3.62 | 3.64 | 3.66 | 3.68 | 3.71 | 3.74 | 3.78 | 3.82 | 3.86 |
| 38 | 3.48 | 3.48 | 3.49 | 3.50 | 3.51 | 3.52 | 3.53 | 3.55 | 3.56 | 3.57 | 3.59 | 3.61 | 3.63 | 3.65 | 3.67 | 3.69 | 3.72 | 3.75 | 3.79 | 3.83 |
| 39 | 3.48 | 3.48 | 3.49 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.56 | 3.57 | 3.58 | 3.60 | 3.62 | 3.64 | 3.66 | 3.68 | 3.70 | 3.73 | 3.76 | 3.80 |
| 40 | 3.48 | 3.49 | 3.49 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.55 | 3.57 | 3.58 | 3.59 | 3.61 | 3.63 | 3.65 | 3.67 | 3.69 | 3.72 | 3.74 | 3.78 |
| 41 | 3.48 | 3.49 | 3.49 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.55 | 3.56 | 3.58 | 3.59 | 3.60 | 3.62 | 3.64 | 3.66 | 3.68 | 3.70 | 3.73 | 3.75 |
| 42 | 3.48 | 3.49 | 3.50 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.60 | 3.61 | 3.63 | 3.65 | 3.67 | 3.69 | 3.71 | 3.74 |
| 43 | 3.48 | 3.49 | 3.50 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.59 | 3.61 | 3.62 | 3.64 | 3.66 | 3.68 | 3.70 | 3.72 |
| 44 | 3.49 | 3.49 | 3.50 | 3.50 | 3.51 | 3.52 | 3.53 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.59 | 3.60 | 3.62 | 3.63 | 3.65 | 3.67 | 3.68 | 3.71 |
| 45 | 3.49 | 3.49 | 3.50 | 3.51 | 3.51 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.59 | 3.60 | 3.61 | 3.62 | 3.64 | 3.66 | 3.67 | 3.69 |
| 46 | 3.49 | 3.49 | 3.50 | 3.51 | 3.51 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.59 | 3.61 | 3.62 | 3.63 | 3.65 | 3.66 | 3.68 |
| 47 | 3.49 | 3.50 | 3.50 | 3.51 | 3.51 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.59 | 3.60 | 3.61 | 3.63 | 3.64 | 3.66 | 3.67 |
| 48 | 3.49 | 3.50 | 3.50 | 3.51 | 3.52 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.59 | 3.60 | 3.61 | 3.62 | 3.63 | 3.65 | 3.66 |
| 49 | 3.49 | 3.50 | 3.51 | 3.51 | 3.52 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.58 | 3.59 | 3.61 | 3.62 | 3.63 | 3.64 | 3.66 |
| 50 | 3.50 | 3.50 | 3.51 | 3.51 | 3.52 | 3.52 | 3.53 | 3.54 | 3.54 | 3.55 | 3.56 | 3.57 | 3.58 | 3.58 | 3.59 | 3.60 | 3.61 | 3.62 | 3.64 | 3.65 |

$\alpha = 0.10$

P, the number of parameters

| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 6 | 3.91 | | | | | | | | | | | | | | | | | | | |
| 7 | 3.64 | 4.09 | | | | | | | | | | | | | | | | | | |
| 8 | 3.48 | 3.76 | 4.25 | | | | | | | | | | | | | | | | | |
| 9 | 3.39 | 3.58 | 3.88 | 4.40 | | | | | | | | | | | | | | | | |
| 10 | 3.32 | 3.47 | 3.67 | 3.98 | 4.53 | | | | | | | | | | | | | | | |
| 11 | 3.28 | 3.39 | 3.54 | 3.75 | 4.08 | 4.66 | | | | | | | | | | | | | | |
| 12 | 3.25 | 3.33 | 3.45 | 3.60 | 3.82 | 4.17 | 4.77 | | | | | | | | | | | | | |
| 13 | 3.23 | 3.30 | 3.38 | 3.50 | 3.66 | 3.89 | 4.25 | 4.88 | | | | | | | | | | | | |
| 14 | 3.21 | 3.27 | 3.34 | 3.43 | 3.55 | 3.72 | 3.95 | 4.33 | 4.99 | | | | | | | | | | | |
| 15 | 3.20 | 3.25 | 3.31 | 3.38 | 3.47 | 3.60 | 3.77 | 4.01 | 4.40 | 5.08 | | | | | | | | | | |
| 16 | 3.19 | 3.23 | 3.28 | 3.34 | 3.42 | 3.52 | 3.64 | 3.82 | 4.07 | 4.47 | 5.17 | | | | | | | | | |
| 17 | 3.18 | 3.22 | 3.26 | 3.31 | 3.38 | 3.46 | 3.55 | 3.68 | 3.86 | 4.12 | 4.53 | 5.26 | | | | | | | | |
| 18 | 3.18 | 3.21 | 3.25 | 3.29 | 3.35 | 3.41 | 3.49 | 3.59 | 3.72 | 3.91 | 4.18 | 4.60 | 5.34 | | | | | | | |
| 19 | 3.18 | 3.20 | 3.24 | 3.28 | 3.32 | 3.37 | 3.44 | 3.52 | 3.63 | 3.76 | 3.95 | 4.22 | 4.66 | 5.42 | | | | | | |
| 20 | 3.17 | 3.20 | 3.23 | 3.26 | 3.30 | 3.35 | 3.40 | 3.47 | 3.55 | 3.66 | 3.80 | 3.99 | 4.27 | 4.71 | 5.50 | | | | | |
| 21 | 3.17 | 3.20 | 3.22 | 3.25 | 3.29 | 3.33 | 3.37 | 3.43 | 3.50 | 3.58 | 3.69 | 3.83 | 4.03 | 4.31 | 4.77 | 5.57 | | | | |
| 22 | 3.17 | 3.19 | 3.22 | 3.24 | 3.27 | 3.31 | 3.35 | 3.40 | 3.45 | 3.52 | 3.61 | 3.72 | 3.86 | 4.06 | 4.36 | 4.82 | 5.65 | | | |
| 23 | 3.17 | 3.19 | 3.21 | 3.24 | 3.26 | 3.30 | 3.33 | 3.37 | 3.42 | 3.48 | 3.55 | 3.64 | 3.75 | 3.90 | 4.10 | 4.40 | 4.87 | 5.71 | | |
| 24 | 3.18 | 3.19 | 3.21 | 3.23 | 3.26 | 3.28 | 3.32 | 3.35 | 3.39 | 3.44 | 3.50 | 3.57 | 3.66 | 3.78 | 3.93 | 4.13 | 4.44 | 4.92 | 5.78 | |
| 25 | 3.18 | 3.19 | 3.21 | 3.23 | 3.25 | 3.28 | 3.30 | 3.33 | 3.37 | 3.41 | 3.46 | 3.52 | 3.60 | 3.69 | 3.80 | 3.96 | 4.17 | 4.47 | 4.96 | 5.84 |
| 26 | 3.18 | 3.19 | 3.21 | 3.23 | 3.25 | 3.27 | 3.29 | 3.32 | 3.35 | 3.39 | 3.43 | 3.48 | 3.54 | 3.62 | 3.71 | 3.83 | 3.98 | 4.20 | 4.51 | 5.01 |
| 27 | 3.18 | 3.19 | 3.21 | 3.23 | 3.24 | 3.26 | 3.29 | 3.31 | 3.34 | 3.37 | 3.41 | 3.45 | 3.50 | 3.56 | 3.64 | 3.73 | 3.85 | 4.01 | 4.23 | 4.55 |
| 28 | 3.18 | 3.20 | 3.21 | 3.22 | 3.24 | 3.26 | 3.28 | 3.30 | 3.33 | 3.36 | 3.39 | 3.43 | 3.47 | 3.52 | 3.58 | 3.66 | 3.75 | 3.88 | 4.04 | 4.26 |
| 29 | 3.19 | 3.20 | 3.21 | 3.22 | 3.24 | 3.26 | 3.27 | 3.29 | 3.32 | 3.34 | 3.37 | 3.40 | 3.44 | 3.49 | 3.54 | 3.60 | 3.68 | 3.78 | 3.90 | 4.06 |
| 30 | 3.19 | 3.20 | 3.21 | 3.22 | 3.24 | 3.25 | 3.27 | 3.29 | 3.31 | 3.33 | 3.36 | 3.39 | 3.42 | 3.46 | 3.50 | 3.56 | 3.62 | 3.70 | 3.80 | 3.92 |
| 31 | 3.19 | 3.20 | 3.21 | 3.22 | 3.24 | 3.25 | 3.27 | 3.28 | 3.30 | 3.32 | 3.35 | 3.37 | 3.40 | 3.44 | 3.48 | 3.52 | 3.58 | 3.64 | 3.72 | 3.82 |
| 32 | 3.19 | 3.20 | 3.21 | 3.23 | 3.24 | 3.25 | 3.26 | 3.28 | 3.30 | 3.32 | 3.34 | 3.36 | 3.39 | 3.42 | 3.45 | 3.49 | 3.54 | 3.59 | 3.66 | 3.74 |
| 33 | 3.20 | 3.21 | 3.22 | 3.23 | 3.24 | 3.25 | 3.26 | 3.28 | 3.29 | 3.31 | 3.33 | 3.35 | 3.37 | 3.40 | 3.43 | 3.47 | 3.51 | 3.55 | 3.61 | 3.67 |
| 34 | 3.20 | 3.21 | 3.22 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.29 | 3.31 | 3.32 | 3.34 | 3.36 | 3.39 | 3.41 | 3.45 | 3.48 | 3.52 | 3.57 | 3.62 |
| 35 | 3.20 | 3.21 | 3.22 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.29 | 3.30 | 3.32 | 3.33 | 3.35 | 3.38 | 3.40 | 3.43 | 3.46 | 3.49 | 3.53 | 3.58 |
| 36 | 3.21 | 3.21 | 3.22 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.28 | 3.30 | 3.31 | 3.33 | 3.35 | 3.37 | 3.39 | 3.41 | 3.44 | 3.47 | 3.51 | 3.55 |
| 37 | 3.21 | 3.22 | 3.22 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.28 | 3.30 | 3.31 | 3.32 | 3.34 | 3.36 | 3.38 | 3.40 | 3.42 | 3.45 | 3.48 | 3.52 |
| 38 | 3.21 | 3.22 | 3.23 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.28 | 3.29 | 3.31 | 3.32 | 3.33 | 3.35 | 3.37 | 3.39 | 3.41 | 3.44 | 3.46 | 3.50 |
| 39 | 3.22 | 3.22 | 3.23 | 3.24 | 3.24 | 3.25 | 3.26 | 3.27 | 3.28 | 3.29 | 3.30 | 3.32 | 3.33 | 3.35 | 3.36 | 3.38 | 3.40 | 3.42 | 3.45 | 3.48 |
| 40 | 3.22 | 3.23 | 3.23 | 3.24 | 3.25 | 3.25 | 3.26 | 3.27 | 3.28 | 3.29 | 3.30 | 3.31 | 3.33 | 3.34 | 3.36 | 3.37 | 3.39 | 3.41 | 3.43 | 3.46 |
| 41 | 3.22 | 3.23 | 3.23 | 3.24 | 3.25 | 3.26 | 3.26 | 3.27 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.34 | 3.35 | 3.37 | 3.38 | 3.40 | 3.42 | 3.44 |
| 42 | 3.23 | 3.23 | 3.24 | 3.24 | 3.25 | 3.26 | 3.26 | 3.27 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.33 | 3.35 | 3.36 | 3.38 | 3.39 | 3.41 | 3.43 |
| 43 | 3.23 | 3.23 | 3.24 | 3.25 | 3.25 | 3.26 | 3.27 | 3.27 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.33 | 3.34 | 3.36 | 3.37 | 3.39 | 3.40 | 3.42 |
| 44 | 3.23 | 3.24 | 3.24 | 3.25 | 3.25 | 3.26 | 3.27 | 3.27 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.38 | 3.39 | 3.41 |
| 45 | 3.24 | 3.24 | 3.25 | 3.25 | 3.26 | 3.26 | 3.27 | 3.28 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.37 | 3.39 | 3.40 |
| 46 | 3.24 | 3.24 | 3.25 | 3.25 | 3.26 | 3.26 | 3.27 | 3.28 | 3.28 | 3.29 | 3.30 | 3.31 | 3.32 | 3.32 | 3.33 | 3.34 | 3.36 | 3.37 | 3.38 | 3.40 |
| 47 | 3.24 | 3.25 | 3.25 | 3.26 | 3.26 | 3.27 | 3.27 | 3.28 | 3.28 | 3.29 | 3.30 | 3.31 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.38 | 3.39 |
| 48 | 3.25 | 3.25 | 3.25 | 3.26 | 3.26 | 3.27 | 3.27 | 3.28 | 3.29 | 3.29 | 3.30 | 3.31 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.37 | 3.39 |
| 49 | 3.25 | 3.25 | 3.26 | 3.26 | 3.27 | 3.27 | 3.28 | 3.28 | 3.29 | 3.29 | 3.30 | 3.31 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.37 | 3.38 |
| 50 | 3.25 | 3.26 | 3.26 | 3.26 | 3.27 | 3.27 | 3.28 | 3.28 | 3.29 | 3.29 | 3.30 | 3.31 | 3.31 | 3.32 | 3.33 | 3.34 | 3.35 | 3.36 | 3.37 | 3.38 |

(Table A28 continued on next page.)

TABLE A28 (Cont'd)

| $\alpha = 0.20$ | | | | | | | | | | | | | | | | | | | | |
|-----------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| P, the number of parameters | | | | | | | | | | | | | | | | | | | | |
| N | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 6 | 3.09 | | | | | | | | | | | | | | | | | | | |
| 7 | 2.96 | 3.25 | | | | | | | | | | | | | | | | | | |
| 8 | 2.90 | 3.08 | 3.39 | | | | | | | | | | | | | | | | | |
| 9 | 2.86 | 2.99 | 3.18 | 3.52 | | | | | | | | | | | | | | | | |
| 10 | 2.83 | 2.93 | 3.07 | 3.28 | 3.63 | | | | | | | | | | | | | | | |
| 11 | 2.82 | 2.89 | 3.00 | 3.14 | 3.36 | 3.74 | | | | | | | | | | | | | | |
| 12 | 2.81 | 2.87 | 2.95 | 3.06 | 3.21 | 3.44 | 3.84 | | | | | | | | | | | | | |
| 13 | 2.81 | 2.86 | 2.92 | 3.00 | 3.11 | 3.27 | 3.51 | 3.93 | | | | | | | | | | | | |
| 14 | 2.81 | 2.85 | 2.90 | 2.97 | 3.05 | 3.17 | 3.33 | 3.58 | 4.02 | | | | | | | | | | | |
| 15 | 2.81 | 2.85 | 2.89 | 2.94 | 3.01 | 3.10 | 3.21 | 3.39 | 3.65 | 4.10 | | | | | | | | | | |
| 16 | 2.81 | 2.84 | 2.88 | 2.92 | 2.98 | 3.05 | 3.14 | 3.26 | 3.44 | 3.71 | 4.18 | | | | | | | | | |
| 17 | 2.82 | 2.84 | 2.87 | 2.91 | 2.96 | 3.01 | 3.09 | 3.18 | 3.30 | 3.49 | 3.77 | 4.26 | | | | | | | | |
| 18 | 2.82 | 2.84 | 2.87 | 2.90 | 2.94 | 2.99 | 3.05 | 3.12 | 3.22 | 3.35 | 3.53 | 3.82 | 4.33 | | | | | | | |
| 19 | 2.83 | 2.85 | 2.87 | 2.90 | 2.93 | 2.97 | 3.02 | 3.08 | 3.15 | 3.25 | 3.39 | 3.58 | 3.88 | 4.39 | | | | | | |
| 20 | 2.83 | 2.85 | 2.87 | 2.90 | 2.92 | 2.96 | 3.00 | 3.05 | 3.11 | 3.19 | 3.29 | 3.42 | 3.62 | 3.93 | 4.46 | | | | | |
| 21 | 2.84 | 2.85 | 2.87 | 2.89 | 2.92 | 2.95 | 2.98 | 3.03 | 3.08 | 3.14 | 3.22 | 3.32 | 3.46 | 3.66 | 3.97 | 4.52 | | | | |
| 22 | 2.84 | 2.86 | 2.87 | 2.89 | 2.92 | 2.94 | 2.97 | 3.01 | 3.05 | 3.10 | 3.16 | 3.24 | 3.35 | 3.49 | 3.70 | 4.02 | 4.58 | | | |
| 23 | 2.85 | 2.86 | 2.88 | 2.90 | 2.92 | 2.94 | 2.96 | 3.00 | 3.03 | 3.07 | 3.13 | 3.19 | 3.27 | 3.38 | 3.53 | 3.74 | 4.07 | 4.64 | | |
| 24 | 2.85 | 2.87 | 2.88 | 2.90 | 2.91 | 2.94 | 2.96 | 2.99 | 3.02 | 3.05 | 3.10 | 3.15 | 3.22 | 3.30 | 3.41 | 3.56 | 3.77 | 4.11 | 4.70 | |
| 25 | 2.86 | 2.87 | 2.88 | 2.90 | 2.92 | 2.93 | 2.95 | 2.98 | 3.01 | 3.04 | 3.07 | 3.12 | 3.17 | 3.24 | 3.32 | 3.44 | 3.59 | 3.81 | 4.15 | 4.75 |
| 26 | 2.86 | 2.88 | 2.89 | 2.90 | 2.92 | 2.93 | 2.95 | 2.97 | 3.00 | 3.02 | 3.06 | 3.09 | 3.14 | 3.19 | 3.26 | 3.35 | 3.46 | 3.62 | 3.84 | 4.19 |
| 27 | 2.87 | 2.88 | 2.89 | 2.90 | 2.92 | 2.93 | 2.95 | 2.97 | 2.99 | 3.01 | 3.04 | 3.08 | 3.11 | 3.16 | 3.22 | 3.28 | 3.37 | 3.49 | 3.65 | 3.87 |
| 28 | 2.88 | 2.89 | 2.90 | 2.91 | 2.92 | 2.93 | 2.95 | 2.97 | 2.99 | 3.01 | 3.03 | 3.06 | 3.09 | 3.13 | 3.18 | 3.24 | 3.31 | 3.40 | 3.51 | 3.67 |
| 29 | 2.88 | 2.89 | 2.90 | 2.91 | 2.92 | 2.94 | 2.95 | 2.96 | 2.98 | 3.00 | 3.02 | 3.05 | 3.08 | 3.11 | 3.15 | 3.20 | 3.26 | 3.33 | 3.42 | 3.54 |
| 30 | 2.89 | 2.90 | 2.91 | 2.91 | 2.93 | 2.94 | 2.95 | 2.96 | 2.98 | 3.00 | 3.02 | 3.04 | 3.07 | 3.09 | 3.13 | 3.17 | 3.22 | 3.27 | 3.35 | 3.44 |
| 31 | 2.89 | 2.90 | 2.91 | 2.92 | 2.93 | 2.94 | 2.95 | 2.96 | 2.98 | 2.99 | 3.01 | 3.03 | 3.05 | 3.08 | 3.11 | 3.14 | 3.19 | 3.23 | 3.29 | 3.37 |
| 32 | 2.90 | 2.91 | 2.91 | 2.92 | 2.93 | 2.94 | 2.95 | 2.96 | 2.98 | 2.99 | 3.01 | 3.03 | 3.05 | 3.07 | 3.10 | 3.13 | 3.16 | 3.20 | 3.25 | 3.31 |
| 33 | 2.90 | 2.91 | 2.92 | 2.93 | 2.94 | 2.94 | 2.95 | 2.97 | 2.98 | 2.99 | 3.01 | 3.02 | 3.04 | 3.06 | 3.08 | 3.11 | 3.14 | 3.18 | 3.22 | 3.27 |
| 34 | 2.91 | 2.92 | 2.92 | 2.93 | 2.94 | 2.95 | 2.96 | 2.97 | 2.98 | 2.99 | 3.00 | 3.02 | 3.04 | 3.05 | 3.07 | 3.10 | 3.12 | 3.16 | 3.19 | 3.23 |
| 35 | 2.91 | 2.92 | 2.93 | 2.93 | 2.94 | 2.95 | 2.96 | 2.97 | 2.98 | 2.99 | 3.00 | 3.02 | 3.03 | 3.05 | 3.07 | 3.09 | 3.11 | 3.14 | 3.17 | 3.21 |
| 36 | 2.92 | 2.93 | 2.93 | 2.94 | 2.95 | 2.95 | 2.96 | 2.97 | 2.98 | 2.99 | 3.00 | 3.01 | 3.03 | 3.04 | 3.06 | 3.08 | 3.10 | 3.12 | 3.15 | 3.18 |
| 37 | 2.93 | 2.93 | 2.94 | 2.94 | 2.95 | 2.96 | 2.96 | 2.97 | 2.98 | 2.99 | 3.00 | 3.01 | 3.03 | 3.04 | 3.06 | 3.07 | 3.09 | 3.11 | 3.14 | 3.16 |
| 38 | 2.93 | 2.94 | 2.94 | 2.95 | 2.95 | 2.96 | 2.97 | 2.98 | 2.98 | 2.99 | 3.00 | 3.01 | 3.02 | 3.04 | 3.05 | 3.07 | 3.08 | 3.10 | 3.13 | 3.15 |
| 39 | 2.94 | 2.94 | 2.95 | 2.95 | 2.96 | 2.96 | 2.97 | 2.98 | 2.99 | 2.99 | 3.00 | 3.01 | 3.02 | 3.04 | 3.05 | 3.06 | 3.08 | 3.10 | 3.12 | 3.14 |
| 40 | 2.94 | 2.94 | 2.95 | 2.96 | 2.96 | 2.97 | 2.97 | 2.98 | 2.99 | 3.00 | 3.00 | 3.01 | 3.02 | 3.03 | 3.05 | 3.06 | 3.07 | 3.09 | 3.11 | 3.13 |
| 41 | 2.94 | 2.95 | 2.95 | 2.96 | 2.96 | 2.97 | 2.98 | 2.98 | 2.99 | 3.00 | 3.01 | 3.01 | 3.02 | 3.03 | 3.04 | 3.06 | 3.07 | 3.08 | 3.10 | 3.12 |
| 42 | 2.95 | 2.95 | 2.96 | 2.96 | 2.97 | 2.97 | 2.98 | 2.99 | 2.99 | 3.00 | 3.01 | 3.02 | 3.02 | 3.03 | 3.04 | 3.05 | 3.07 | 3.08 | 3.09 | 3.11 |
| 43 | 2.95 | 2.96 | 2.96 | 2.97 | 2.97 | 2.98 | 2.98 | 2.99 | 3.00 | 3.00 | 3.01 | 3.02 | 3.02 | 3.03 | 3.04 | 3.05 | 3.06 | 3.08 | 3.09 | 3.10 |
| 44 | 2.96 | 2.96 | 2.97 | 2.97 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.01 | 3.02 | 3.03 | 3.03 | 3.04 | 3.05 | 3.06 | 3.07 | 3.09 | 3.10 |
| 45 | 2.96 | 2.97 | 2.97 | 2.98 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.01 | 3.01 | 3.02 | 3.03 | 3.04 | 3.05 | 3.06 | 3.07 | 3.08 | 3.10 |
| 46 | 2.97 | 2.97 | 2.98 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.01 | 3.02 | 3.02 | 3.03 | 3.04 | 3.05 | 3.06 | 3.07 | 3.08 | 3.09 | 3.10 |
| 47 | 2.97 | 2.98 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.01 | 3.01 | 3.02 | 3.02 | 3.03 | 3.04 | 3.04 | 3.05 | 3.06 | 3.07 | 3.08 | 3.09 |
| 48 | 2.98 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.00 | 3.01 | 3.01 | 3.02 | 3.03 | 3.03 | 3.04 | 3.04 | 3.05 | 3.06 | 3.07 | 3.08 | 3.09 |
| 49 | 2.98 | 2.98 | 2.99 | 2.99 | 3.00 | 3.00 | 3.00 | 3.01 | 3.01 | 3.02 | 3.02 | 3.03 | 3.03 | 3.04 | 3.05 | 3.05 | 3.06 | 3.07 | 3.08 | 3.09 |
| 50 | 2.99 | 2.99 | 2.99 | 3.00 | 3.00 | 3.00 | 3.01 | 3.01 | 3.02 | 3.02 | 3.03 | 3.03 | 3.04 | 3.04 | 3.05 | 3.05 | 3.06 | 3.07 | 3.08 | 3.09 |

*The values were computed by Kirkland B. Stewart in work performed at the Battelle Pacific Northwest Laboratory under Department of Energy Contract DE-AC06-76RLO-1830.

INDEX

- abrupt loss, 863
- acceptance number, 884
- accounting adjustments for
 - bias corrections, 628, 645
 - bookkeeping mistakes, 628, 645
 - change of enrichment, 628, 646
 - change of ownership, 628
 - decay, fission or transmutation, 628, 646
 - inventory difference, 628, 644-645
 - measurement mistakes, 628, 645
 - shipper-receiver differences, 628, 644-645
- accounting data, verification of, 871
- accounting errors, effect of, 832
- accounting for holdup, 644
- accounting for misplaced material, 644
- accounting process, 669
- accounting records, 639-643
- accounting results
 - evaluation of, 831
 - expected values of, 683
 - variances of, 683
- accounting structure, 669
- accounting systems, 627, 639
- adjustments, *see* accounting adjustments
- admissible strategy, 610
- alarm limits, 788
- aliased effects, 420
- alias structure, 422
- alternative hypothesis, 136
- analysis of variance (ANOVA), 233
 - nested mixed effects, 299
 - nested random effects, 283
 - one-way fixed effects, 234
 - one-way random effects, 264
 - two-way crossed classifications, 310
 - two-way fixed effects, 313
 - two-way mixed effects, 324
 - two-way random effects, 320
- analytical report, 641
- ANOVA (analysis of variance), 233
- approximate degrees of freedom,
see Satterthwaite approximation
- approximate value of an integral, 580
- arithmetic mean, 5
- assignable cause, 210
- attribute sampling, 880
 - binomial distribution, 888
 - approximation to the hypergeometric, 889
 - fraction sampled, 891
 - required sample size, 891
 - type II error probability, 891
 - hypergeometric distribution, 881
 - acceptance number, 884
 - fractional number defective, 888
 - fraction defective, 881
 - number defective, 881
 - probability of detection, 881
 - required sample size, 882, 887
 - Type II error probability, 887
 - zero acceptance number, 886
- attributes measurement, 875
- auditing, 871
 - capability, 873
 - independent assurance, 873
 - performance, 873
- audit process, 871
- available actions, 603
- average, 5
- average sample number (ASN), 203, 209
- bar chart, 12
- Bartlett's test for equal variances, 802
- basic measurement result, 669
- Bayesian estimators, 102
- Bayes risk, 609, 614
- Bayes strategy, 609, 613
- Behrens-Fisher problem, 168
- bell-shaped curve, 57

- beta distribution, 69
 - pdf, parameters, mean, variance, 68
- bias, 672
 - accounting adjustments for, 645
 - correction, 722, 785, 805
- biased estimator, 98
- binomial approximation to the hypergeometric, 49, 889
- binomial distribution, 47, 888
 - estimator for parameter p , 106
 - normal approximation to, 62
 - pdf, mean, variance, 47
 - Poisson approximation to, 51
 - sample proportion, 106
 - table, 928
- bivariate distribution, continuous, 73
 - conditional mean and variance, 76
 - conditional pdf, 74
 - correlation coefficient, 75
 - covariance, 75
 - independence, 75, 76
 - joint cdf, 74
 - joint pdf, 74
 - marginal pdf, 74
 - mean, variance, 75
- bivariate distribution, discrete, 70
 - conditional mean and variance, 73
 - conditional pdf, 71
 - correlation coefficient, 73
 - covariance, 73
 - independence, 73
 - joint cdf, 71
 - joint pdf, 70
 - marginal pdf, 71
 - mean, variance, 72
- bivariate normal distribution, 82
 - conditional mean, 85
 - conditional pdf, 84
 - conditional variance, 85
 - correlation coefficient, 83
 - joint pdf, parameters, 83
 - marginal pdf, 83, 84
 - means, variances, covariance, 83
 - regression, 85
- blocking with fractional replicates, 427
- Bonferroni
 - inequality, 539
 - intervals, 251
 - t statistics, 543
- book inventory, 620, 639
- boxplot, 17
- calibration
 - general linear, 737
- calibration (cont'd)
 - parameter estimates, 738
 - weighted regression analysis, 748
- model, 733
 - nonlinear, 760
 - of measurement systems, 732
 - of NDA instruments, 732
 - of process vessels, 774
 - process of, 732
 - single point, 732
 - uncertainty, 763
- calibration of measurement processes, 633
- cdf (cumulative distribution function), 40
- center line, 210
- Central Limit Theorem, 61
- central tendency, 4
- chance event, 30
- Chebyshev's inequality, 214
- chemical analyses, 653-657
 - see also* measurement methods
- chi-square distribution, 69
 - pdf, parameters, mean, variance, 68, 69, 108
 - table, 937
- chi-square sum, 544
- chi-square test, 523
- classification of results by error source, 680
- class mark, 5
- class mean, 235
- cluster sampling, one-stage, 455
 - estimators for mean and total, 455
 - required sample size, 456
 - standard error of estimators, 455
- cluster sampling, two-stage 458
 - estimators for mean and total, 459
 - required sample size, 459
 - standard error of estimators, 459
- coefficient of determination, R^2 , 349
- combination, 32
- combination of variance estimators, 699
- common error effects in ID components, 700
- complement of a set, 30
- completely randomized design, 391
- compliance tests, accounting data, 873
- components of a material balance, 629-630
- compound smoothing, 22
- computerized model, 570
- concordant pairs, 521
- conditional joint pdf, 77
- conditional mean, 73, 76, 85
- conditional pdf, 71, 74, 84
- conditional probability, 35
- conditional variance, 73, 76, 85

- confidence interval, 96, 106
 for binomial distribution parameter p ,
 118-120
 exact, 120
 normal approximation, 118
 required sample size, 119
 for difference of two binomial parameters,
 121
 for difference of two means, 115
 variances known, 116
 variances unknown but equal, 116
 variances unknown and unequal, 116
 for finite population mean and total, 481
 for hypergeometric distribution
 parameter D , 122
 for mean of a linear combination, 127
 for mean of normal distribution, 113
 required sample size, 114
 variance known, 113
 variance unknown, 113
 for ratio of two standard deviations,
 111
 for ratio of two variances, 111
 for standard deviation, 109
 confidence intervals and hypothesis testing,
 151
 confounded effects, 420
 congruential generator, 577
 consistent estimator, 98
 constant absolute variance, 673
 constant bias effect, 673
 constant error effect, 673
 constant loss, 864
 constant relative variance, 673
 constrained maximum likelihood estimate,
 102
 contingency table, 502
 continuous model, 573
 continuous monitoring, 790
 continuous random variable, 1, 38
 contrasts, 253, 305
 control chart, 210, 787, 851
 Cook's distance, 380
 corrected inventory difference, 909
 variance of, 909
 correcting measurement results, 722
 correlated statements, 541
 correlation, 181, 332
 correlation coefficient, 73, 75, 79, 83,
 332-333
 population correlation coefficient, 332
 sample correlation coefficient, 333
 covariance, 73, 75, 78, 79
 covariance matrix, *see* variance-
 covariance matrix
 critical region, 137
 cumulative distribution function (cdf), 40
 cumulative error model, 774
 parameter estimates, 776
 problems with, 777
 cumulative inventory difference, 831
 cumulative sum (CUSUM) chart, 224
 amount of shift, 227
 lead distance, 228
 turning point, 227
 Type I and Type II errors, 227
 V-mask, 224
 curtailed sequential sampling, 443
 CUSUM (cumulative sum) chart, 206,
 224
 data smoothing, 22
 data splitting in regression, 382
 decile, 11
 decision analysis, 602
 decision rule, selection of, 859
 decision theory, 601
 defining contrast, 422
 dependent error effects, 679
 dependent variable, 389
 design of experiments, 387
 deterministic model, 572
 difference statistic, 903
 variance of, 906
 discordant pairs, 521
 discrete model, 573
 discrete random variable, 1, 38
 dispersion, 9
 distribution of sample mean, 112
 distribution, continuous, 53
 beta distribution, 69
 bivariate normal distribution, 82
 chi-square distribution, 69
 exponential distribution, 67
 F-distribution, 110
 gamma distribution, 67
 lognormal distribution, 64
 normal distribution, 56
 standard normal distribution, 58
 uniform distribution, 53
 Weibull distribution, 69
 distribution, discrete, 44
 binomial distribution, 47
 hypergeometric distribution, 44
 multinomial distribution, 79
 Poisson distribution, 50
 diversion
 abrupt, 850, 863
 protracted, 850
 Duncan's multiple range test, 256, 309

- Duncan's multiple range test (cont'd)
 equal sample sizes, 257
 tables, 951
 unequal sample sizes, 258
- DUPLEX algorithm, 383
- Durbin test, 517
- effect of accounting errors, 832
- elementary event, 30
- empirical cdf, 96, 529
- empirical control chart, 211
 count rate, 222
 mean, 218
 proportion, 221
 range, 223
- empirical density function, 96
- empirical distribution, 96
- error rate, 538
- errors in independent variables, 343
- error size classifications, 876
 gross defect, 876
 medium defect, 876
 significant quantity, 876
 small defect, 876
- error structure, 670
 generic treatment of, 671
 simplification of, 691
 sources of variability, 670
- error variance component, 674
 models for estimating, 674
- error variances, compilation of, 692
- estimate, 95
 maximum likelihood, 101
- estimating means and totals, 447
- estimating measurement variability, 634
- estimation, 95
- estimation of component variances, 700
- estimator, 95
 point, 95
 method of moments, 99
 interval, 95, 106
- event, 30, 33
- expected error rate, 540
- expected loss, 605
- expected mean squares, 239, 266, 285,
 300, 321, 326
see also ANOVA
- expected value, 41
- experiment, 29
- experimental design, 387
- experimental error variance, 414
- experimental objectives, 389
- experimental units, 388
- exploration of a system, 575
- exploratory data analysis, 2
- exponential distribution, 67
 pdf, parameters, mean, variance, 68
- facility and process variables, 625-626
- factorial designs, 403-429
 2^k designs, 409
 3^k designs, 416
- factor, 389
- false alarm rate, 138
- false alarm rate, choice of, 859
- family of statements, 538
- F-distribution, 110
 degrees of freedom, 110
 tables, 940
- feasibility of an experiment, 390
- filter, 22
- finite population correction factor, 50
- finite populations, sampling, 439
- Fisher's combination of probabilities test,
 546
- Fisher-Cornish approximation, 561
- fractional factorial designs, 419
 2^{k-p} designs, 421
 3^{k-p} designs, 427
- fractions of:
 2^k designs, 421
 3^k designs, 427
 general factorial designs, 427
- Freund-Ansari test, 500
- Friedman test, 514
- F-test for two variances, 171, 501
- functions of random variables, 88
 general functions, 90
 approximate mean and variance, 91
 estimator for approximate mean, 126
 estimator for approximate variance, 126
 Taylor series approximation, 91
 linear combinations, 88
 estimator for mean, 124
 estimator for variance, 124
 mean and variance of, 89
- game theory, 601
- gamma distribution, 67
 parameter estimators, 107
 pdf, parameters, mean, variance, 68
- general functions of variables, 90
 approximate mean and variance, 91
 estimate for approximate mean, 126
 estimator for approximate variance, 126
 Taylor series approximation, 91
- general linear calibration, 737

- general linear regression model, 344
 generating random deviates, 576
 see also random deviate generation
 geometric mean, 8
 geometric standard deviation, 10
 goodness of fit tests, 523
 chi-square test, 523
 Kolmogorov-Smirnov test, 529
 Kolmogorov test, 529
 Lilliefors test for normality, 531
 Shapiro-Wilk test for normality, 532
 Graeco-Latin square design, 403
 grouped data, 5
 Grubbs' procedures, 556, 823
- hanning, 26
 hat matrix, 561
 heterogeneous error variances, 343
 hierarchical designs, 433
 hierarchical models ANOVA, 282
 high-level neutron coincidence counter (HLNCC), 748
 calibration of, 751
 high leverage observation, 378
 hinges, 17
 histogram, 12
 HLNCC (high-level neutron coincidence counter), 748
 Hotelling's T^2 -test, 547
 hypergeometric distribution, 44, 881
 binomial approximation to, 49
 parameter estimator, 107
 pdf, mean variance, 44
 Poisson approximation to, 51
 hypothesis testing, 135
 hypothesis testing with confidence intervals, 151
 hypothesis tests for
 binomial parameter p , 183
 approximate test, 188
 OC and power curves, 189
 required sample size, 190
 exact test, 183
 OC and power curves, 185
 required sample size, 186
 comparing two means, variances known, 161
 OC and power curves, 164
 required sample size, 164
 comparing two means, variances unknown, 165
 paired observations, 176
 OC and power curves, 179
- hypothesis tests for (cont'd)
 rationale for pairing, 180
 required sample size, 179
 variances assumed equal, 165
 pooled standard deviation, 167
 pooled variance estimator, 166
 probability of Type II error, 168
 required sample size, 168
 variances unequal, 168
 OC and power curves, 171
 required sample size, 171
 Satterthwaite procedure, 168
 comparing two proportions, 191
 OC and power curves, 193
 required sample size, 193
 comparing two variances, 171
 OC curve, 173
 required sample size, 173
 cumulative inventory differences, 851
 inventory differences, 851
 mean, variance known, 139
 OC and power curves, 144
 one-sided tests, 142
 required sample size, 147-148
 two-sided test, 139
 mean, variance unknown, 152
 OC and power curves, 155
 required sample size, 155
 Poisson parameter, 195
 approximate test, 199
 OC and power curves, 200
 required sample size, 200
 exact test, 195
 OC and power curves, 197
 required sample size, 197
 variance of a normal distribution, 157
 OC and power curves, 159
 required sample size, 160
- ID, *see* inventory difference
 ID components, common error effects, 700
 imperfect inspection, 440, 448, 452
 independence, 73, 75, 76, 79
 independent samples, 161, 495, 501
 independent statements, 541
 independent variables, 389
 indicator variable, 37
 inferences from material accounting, 636
 infinite variance, 734
 influential data points, 379
 inspector's inventory difference, 909
 interaction effect, 404, 413
 interlaboratory exchange studies, 799, 807

internal control, 628
 internal control system, 872
 internal transfer report, 641
 interquartile range, 11
 intersection of two sets, 31
 interval estimator, 95, 106
 interval scale, 4
 intraclass correlation coefficient, 275
 inventory accounting, 621
 inventory difference (ID), 620, 701, 831

- components of, 833
- control chart, 851
- corrected, 909
- redefinition of, 841
- sequences of, 833
- testing, 850, 859

 inventory difference account, 644
 inventory of materials, 620, 643
 inventory strata, 629-630
 inverse calibration, 779
 isolation of effects, 389
 item accounting, 629, 643
 item amounts, 675

- errors in, 675
- total bias effect, 676

 item control areas, 625

Johnson approximation, 97
 joint cdf, 71, 74, 76, 77
 jointly distributed random variables, 76
 joint marginal pdf, 77
 joint pdf, 70, 74, 76, 77

Kendall's tau, 521
 key measurement point, 625
 Kolmogorov-Smirnov test, 529
 Kolmogorov test, 529
 Kruskal-Wallis test, 503
 k-sample squared ranks test, 507

- multiple comparisons, 509

 kurtosis, 43

lack of fit, 349, 356-360

- test for significance of, 359

 Latin square design, 398
 law of large numbers, 35
 lead distance, 228
 least favorable distribution, 614
 least significant range, 257
 least squares, 339
 least squares estimates, 100
 ledgers, 639
 likelihood function, 79, 100

Lilliefors test for normality, 531
 linear combinations of variables, 88

- estimator for mean, 124
- estimator for variance, 124
- mean and variance of, 89

 location, 5, 489, 501
 location parameter, 43
 lognormal distribution, 64

- geometric mean, 66
- mean, median, mode, variance, 65
- parameter estimators, 107
- pdf, parameters, 65

 long-term systematic effect, 692
 loss

- abrupt, 863
- constant, 864
- proportional to throughput, 865
- sequential tests, 866

 losses

- behavior of, 850
- distribution of, 850, 907
- models for, 832, 849, 861
- probability of detecting, 853
 - dependent on loss model, 858
- sampling of, 909
- specialized estimators, 861

 loss function, 603
 loss model, 832, 849, 861
 loss table, 603
 lower control limit (LCL), 210

main effect, 404, 413
 Mann-Whitney test, 495
 manufacturing process, diagram of

- typical, 625, 626

 marginal pdf, 71, 74, 83, 84
 matched pairs tests, 493
 material accounting, 619-622

- inferences from, 636
- requirements, 627
- role of statistics in, 630

 material balance

- areas, 625
- components, 629-630
- equation, 620

 material quantities, inferences about, 903
 material unaccounted for (MUF), 620
 mathematical models, 571
 maximum likelihood estimate, 101
 mean, 42

- mean square error, 98, 792
 - bias component of, 792
- mean vector, 78, 548

- measurement
 - attributes, 875
 - capability verification, 895
 - performance assessment, 896
 - performance verification, 895
 - variables, 875
- measurement control
 - control programs, 632-633
 - use of reference standards, 634
 - procedures, 780
- measurement differences, inferences from, 894
- measurement error, 672
 - biases, 672
 - constant bias effects, 673
 - constant error effect, 673
 - effects, 669
 - random errors, 672
 - relative bias effect, 673
 - relative error effects, 673
 - sources, of, 679
- measurement error, in sampling, 440, 448, 452
- measurement errors, inferences about, 894
- measurement methods, 629, 647-664
 - reference standards for, 648, 651, 658, 662-663
 - statistical control of, 630-636
- measurement process
 - components, 671
 - generic treatment of, 692
- measurement replication, 635
- measurement scale, 2
- measurements, control of bias in, 634
- measurement system
 - modeling, 665
 - accounting structure, 669
 - definitions, 666
 - error structure, 670
 - monitoring of, 767, 780
 - stability of, 752
- measurement uncertainty, 631
 - effect on inventory differences, 631
 - estimation and control, 632-636
- measurement variability
 - controlling, 731
 - data sources for estimating, 731
 - estimating, 731
 - example of modeling, 713
 - modeling, 665
 - sources of, 685
- measures of
 - central tendency, 4
 - influence, 380
- measures of (cont'd)
 - leverage, 380
 - variability, 8
- median, 7
- median test, 501
- method of moments, 99
- minimax criterion, 607
- minimax regret strategy, 615
- minimax risk, 607
- minimax strategy, 607
- minimum variance estimator, 98
- minimum variance unbiased estimator, 98
- mixed level factorial designs, 419
- mixed model, *see* ANOVA
- mixed strategies, 610
- modal group, 8
- mode, 8
- modeling measurement uncertainty, 631
- moment
 - about the mean, 43
 - of a distribution, 43
- Monte Carlo methods, 569
- Monte Carlo simulation, 55, 569
- Mood test, 500
- moving sum chart, 230
- MUF, *see* material unaccounted for
- multinomial distribution, 79
 - mean vector, 81
 - pdf, parameters, 80
 - variance-covariance matrix, 81
- multiple linear regression, 344
 - analysis of variance, 348
 - ordinary least squares estimators, 346
 - weighted least squares estimators, 346
- multistage sampling plan, 441
- multistage sequential sampling, 443
- multivariate analysis, 547
- multivariate analysis of variance, 552
- multivariate confidence region, 552
- multivariate distribution, 76, 548
 - conditional joint pdf, 77
 - conditional mean and variance, 73
 - correlation coefficient, 79
 - covariance, 78, 79
 - independence, 79
 - joint cdf, 76, 77
 - joint marginal pdf, 77
 - joint pdf, 76, 77
 - mean vector, 78
 - variance-covariance matrix, 78
- multivariate statistical methods, 547
- mutually exclusive sets, 32
- nested design, 433

- nested mixed effects ANOVA, 299
 - comparison of means, 305-308
 - contrasts, 305-308
 - confidence intervals, 307-308
 - estimates, 305-306
 - hypothesis tests, 307-308
 - variance of estimates, 306
 - Duncan's multiple range test, 309
 - equal sample sizes, 308-309
 - estimation of means, 301-305
 - estimators, 302-303
 - one-at-a-time confidence intervals, 304
 - simultaneous confidence intervals, 304-305
 - variance of estimators, 302-304
 - weighting schemes, 302
 - expected mean squares, 300
 - mean squares, 300
 - model, 299-300
 - sums of squares, 300
- nested random effects ANOVA, 283
 - equal sample sizes, 298
 - estimating overall mean, 291-298
 - confidence interval, 295
 - estimators, 292-295
 - hypothesis test, 295
 - variance of estimators, 292-295
 - weighting schemes, 292
 - expected mean squares, 285
 - hypothesis tests, 286-287
 - mean squares, 284-285
 - model, 282-283
 - sums of squares, 284
 - variance components, 285-286
- nominal scale, 3
- noncentrality parameter, 155, 167
- noncentral t-distribution, 155, 167
- noninformative prior distribution, 610
- nonlinear calibration, 760
- nonparametric methods, 483
- normal approximation
 - to the binomial, 62, 188
 - to the Poisson, 199
- normal distribution, 56
 - pdf, mean, variance, cdf, 57, 58
 - sample mean, 104
 - sample variance, 105
 - table, 936
- nuclear material accounting, 621-622, 627, 639-646
- nuclear material accounting forms, 627
- null hypothesis, 136
- null set, 30
- objectives of an experiment, 389
- OC (operating characteristic) curve, 144
- one-at-a-time confidence intervals, 249, 304, 317, 328
- one-fourth replicate of a 2^5 design, 425
- one-half replicate of a 2^4 design, 425
- one-way ANOVA, 234, 501
- one-way fixed effects ANOVA, 234
 - comparisons suggested by data, 251
 - all possible contrasts, 262
 - Scheffé's test, 261-262
 - contrasts, 253
 - confidence interval for, 255, 262
 - estimates of, 253
 - hypothesis test for, 254
 - variance of, 254
 - Duncan's multiple range test, 256
 - expected mean squares, 239
 - linear model, 235
 - mean squares, 239
 - normality assumption, 235, 240
 - OC and power curves, 244-249
 - one-at-a-time confidence intervals, 249
 - planned comparisons of means, 253
 - pooled variance estimator, 239
 - required sample sizes, 244-249
 - simultaneous confidence intervals, 250
 - Bonferroni intervals, 251
 - Scheffé intervals, 250
 - sums of squares, 237-239
 - test for equal means, 234
 - unequal sample sizes, 242-244
- one-way random effects ANOVA, 264
 - expected mean squares, 266
 - intraclass correlation coefficient, 275
 - mean squares, 265
 - normality assumption, 267
 - OC and power curves, 270-272
 - overall mean, 274
 - confidence interval for, 279
 - hypothesis tests for, 279
 - minimum variance unbiased estimator, 275
 - sampling schemes, 279-280
 - cost considerations, 280
 - required sample sizes, 280
 - unweighted estimator, 277
 - variance of, 277-278
 - weighted estimator, 276
 - variance of, 276-277
 - random effects model, 265
 - required sample size, 270-273

- one-way random effects ANOVA (cont'd)
 - sums of squares, 265
 - variance components, 265-267
- operating characteristic (OC) curve, 144
- operational activities in simulation, 576
- operator-inspector difference, 898
- optimization of a system, 575
- order statistic, 96, 554
- ordinal scale, 3
- ordinary least squares estimator, 346
- outliers, 537
- outliers in regression, 378, 560
- outliers, testing for, 552

- paired comparisons, 895
- paired observations, 176, 493
- paired t-test, 176, 396, 493
- parameter, 1, 95
- parameter estimation, 95
- pdf (probability density function), 40
- Pearson approximation, 97
- Pearson-Hartley charts, 245, 943
- percentile, 11
- period of a generator, 578
- permutation, 32
- physical inventory, 620, 643-644
- pie chart, 11
- point estimator, 95
- Poisson approximation
 - to the binomial, 51
 - to the hypergeometric, 51
- Poisson distribution, 50
 - parameter estimator, 107
 - pdf, mean, variance, 50, 51
 - table, 934
- polynomial equations, 342
- population, 1
- power curve, 144
- power of a test, 138
- PPS (probability proportional to size)
 - sampling, 443
- PPS cluster sampling without replacement, 470
 - estimator for total, 470
 - required sample size, 471
 - standard error of estimator, 470
- PPS cluster sampling with replacement, 463
 - estimators for mean and total, 463
 - required sample size, 464
 - standard error of estimators, 463

- PPS sampling without replacement, 467
 - estimators for mean and total, 468
 - Rao, Hartley, Cochran (RHC) method, 467
 - required sample size, 469
 - standard error of estimators, 468
- PPS sampling with replacement, 461
 - estimators for mean and total, 461
 - required sample size, 461
 - standard error of estimators, 461
- predicted value, 383-384
 - confidence intervals, 384
 - variance of, 383
- PRESS procedure, 382
- prior distribution, 102, 609
- probability, definition, 29, 33
- probability density function (pdf), 40
- probability distribution, 39
- probability function, 40
- probability model, 39
- probability of detecting defectives, 881
- probability proportional to size (PPS)
 - sampling, 443
- process vessel, calibration of, 774
- pure error, 349
- pure strategy, 610

- Quade test, 510
- quality control of measurement processes, 632-634
- quartile, 11

- R^2 , coefficient of determination, 349
- random deviate generation, 576
 - binomial distribution, 594
 - chi-square distribution, 587
 - F-distribution, 592
 - functions of random variables, 595
 - gamma distribution, 587
 - multivariate normal distribution, 593
 - correlated random deviates, 594
 - noncentral Student's t-distribution, 590
 - normal distribution, 582
 - Box-Muller method, 583
 - Central Limit Theorem, 584
 - polar method, 583
 - Student's t-distribution, 589
 - uniform distribution, 579
- random effect, 692
- random error, 672
- random number generation, 577

- random number tables, 578, 877
 random sample, 1, 441, 577, 877
 random variable, 36
 randomization, 56, 388
 randomized block ANOVA, 395
 randomized block design, 176
 randomized block experiment, 823
 randomized complete block design, 395
 randomized incomplete block design, 403
 randomized strategy, 610
 range, 9
 rank ANOVA test, 516
 rank correlation, 518
 Kendall's tau, 521
 Spearman's rho, 518
 Rao, Hartley, Cochran (RHC) method, 467
 ratio scale, 4
 recalibration procedures, 752
 reference standards in monitoring
 measurement processes, 634
 regression, 85, 331
 regression analysis, 331
 regression coefficient estimators, 346-356
 sampling distribution of, 352
 test for significance of, 355
 variance-covariance matrix, 352-354
 regret function, 615
 rejection region, 137
 related samples, 395, 510
 relative bias effect, 673
 relative error effects, 673
 repeated normal statistics, 542
 replication, 388
 required sample size,
 see ANOVA, hypothesis testing,
 confidence intervals, sampling
 residual sum of squares, 339
 residuals, 27, 339, 362-378
 hat matrix, 371
 plotting residuals, 363
 serial correlation, 373
 Durbin-Watson test, 375
 runs test, 374
 standardized residuals, 370
 test for normality, 369
 resolution III designs, 421
 resolution IV designs, 421
 resolution V designs, 422
 response variable, 389
 robust estimator, 98
 round-off error, 55
 round robin studies, 799
 running average, 26
 running median smoother, 22
 runs test, 484

 S^2 , sample variance, 9
 SAM (stabilized assay meter), 768
 SAM calibration, 768
 sample covariance, 125
 sample mean, 5, 97, 104
 sample mean of grouped data, 5
 sample moment, 97
 sample point, 30, 33
 sample proportion, 106
 sample space, 30, 33
 sample variance of grouped data, 10
 sample variance, S^2 , 9, 105
 sampling, 439
 sampling distribution, 95, 106
 sampling finite populations, 439
 cluster sampling, one-stage, 455
 cluster sampling, two-stage, 458
 confidence intervals for mean and total,
 481
 effect of measurement error, 440, 448,
 452
 PPS cluster sampling without replacement,
 470
 PPS cluster sampling with replacement,
 463
 PPS sampling without replacement, 467
 PPS sampling with replacement, 461
 Rao, Hartley, Cochran (RHC) method,
 467
 sequential sampling, 443
 simple random sampling, 442, 449
 stratified random sampling, 442, 473
 systematic random sampling, 442
 sampling frame, 877
 sampling methods
 for materials, 629, 659-660
 invalid assumptions, 453
 limitations of, 452
 sampling notation, 440
 sampling plan development, 443
 sampling techniques, strengths and
 weaknesses, 444-445
 Satterthwaite approximation, 116, 128,
 168, 699
 see also ANOVA
 scale parameter, 43, 69
 scatter, 9
 scatter plot, 20
 Scheffé's method, 250, 261, 305
 seed value, 578

- sequential material balances, tests for, 866
- sequential probability ratio test (SPRT), 202
 - average sample number (ASN), 203, 209
 - OC curves, 209
 - Type I and Type II errors, 203
- sequential probability ratio test (SPRT) for
 - binomial parameter p , 207
 - average sample number, 209
 - decision regions, 208
 - mean of a normal distribution, 204
 - decision region, 205-207
- sequential sampling, 443
- sequential tests, 202
- serial correlation, 577
- set, 30
- shape parameter, 67
- Shapiro-Wilk test for normality, 532
- Shewhart chart, 212
- shipper-receiver bias, 800
- short-term systematic effect, 692
- short-term systematic error, 679
- Siegel-Tukey test, 500
- significance level, 138
- sign test, 489, 493
- simple effect, 404
- simple linear regression, 336
 - intercept, 338
 - least squares estimates, 340
 - model, 338
 - slope, 338
- simple random sampling, 442, 449
 - confidence intervals, 450
 - estimators for mean and total, 449
 - required sample size, 450
 - standard error of estimators, 449
- simulation analysis, 569
- simulation analysis, advantages of, 570
- simulation model
 - design, 575
 - implementation, 575
 - validation, 575
 - verification, 575
- simultaneous confidence intervals, 250, 304, 542, 544
- simultaneous confidence statement, 544
- simultaneous inference, 537
- simultaneous tests, 550
- single factor experiments, 391
- single point calibration, 732
- skewness, 43
- smoothing, 22
- source data, 639
- sources of error in measurements, 649-651, 655-658, 660-663
 - sources of measurement error, 679
 - sources of measurement variability, 685
 - sources of variability, 670
- SPC (statistical process control), 210
- Spearman's rho, 518
- specialized loss estimators, 861
- split plot design, 430
- SPRT (sequential probability ratio test), 202
- SQC (statistical quality control), 210
- squared ranks test, 498
- stabilized assay meter (SAM), 768
- standard deviation, 10, 43
- standard deviation of sample mean, 104
- standard error, definition, 10
 - see* sampling
- standard error of sample mean, 104
- standard normal distribution, 58
 - mean, variance, cdf, 58
 - table, 936
- standard squares, 400
- standardization of measurement processes, 633
- standardized residuals, 561
- state of nature, 603
- state of statistical control, 210
- state system of accounting and control (SSAC), 632, 872
- statistic, 2, 95
- statistical control, 210
 - of measurement processes, 630
- statistical design of experiments, 387
- statistical hypothesis, 135
- statistical process control (SPC), 210
- statistical quality control (SQC), 210
- statistical sampling, 439
- statistical tables, 927
 - binomial distribution, 928
 - chi-square distribution, 937
 - coefficients for Shapiro-Wilk test, 973
 - critical values for $\sqrt{b_1}$, 980
 - critical values for b_2 , 980
 - critical values for T'_1 and T'_n , 982
 - critical values for T'_n , 981
 - critical values for T_n , 977
 - critical values for w/s , 978
 - Duncan's multiple range test, 951
 - Durbin-Watson test, 955
 - F-distribution, 940
 - Kendall test statistic, 969
 - Kolmogorov test, 971
 - Lilliefors test, 972
 - Mann-Whitney test, 961

- statistical tables (cont'd)
 - normal distribution (standard), 936
 - Pearson-Hartley charts, 943
 - Poisson distribution, 934
 - runs test, 958
 - screening studentized residuals, 983
 - Shapiro-Wilk test statistic, 976
 - Spearman test statistic, 968
 - squared ranks test, 966
 - standardized residuals, 953
 - standard normal distribution, 936
 - Student's t-distribution, 939
 - testing two extreme values, 979
 - Wilcoxon test, 959
- statistically independent, 35
- statistics, role in material accounting, 630
- stem and leaf display, 15
- stepwise regression, 355
- stochastic model, 572
- strategy, 603, 605
- stratified random sampling, 442, 473, 921
 - allocation of sample, 477
 - cluster sampling within strata, 480
 - cost considerations, 474
 - estimators for mean and total, 476
 - PPS sampling within strata, 480
 - required sample size, 478
 - standard error of estimators, 476
 - successful stratification, 473, 476
- stratum totals, 678, 686
 - errors in, 678
 - expected values, 686
 - variances, 686
- studentized residuals, 560
- Student's t-distribution, 113
 - degrees of freedom, 113
 - table, 939
- substantiative tests of accounts, 873
- symbolic models, 571
- systematic error, 632
- systematic random sampling, 442
- system models, 570, 571

- tables of random numbers, 578
- tables, *see* statistical tables
- Taylor series approximation, 91, 763
- t-distribution, *see* Student's t-distribution
- test for high leverage, 380
- test for significance of effects, 415
- test for significance of regression, 350
- test of randomness, 484
 - runs test, 484
- test statistic, 137
- testing all pairs of means, 257
- testing cumulative inventory differences, 850
- testing for outliers, 552
 - Grubbs' treatment of outliers, 556
 - repeated testing for outliers, 558
 - testing largest value, 558
 - testing smallest value, 558
 - testing two extreme values, 558
 - testing two largest values, 558
 - testing two smallest values, 558
- methods of Dixon, 554
 - testing k most extreme values, 556
 - testing largest value, 554
 - testing smallest value, 554
 - testing two extreme values, 555
- testing inventory differences, 850, 859
- tests for comparing several locations, 233, 501
 - independent samples, 234, 501
 - Kruskal-Wallis test, 503
 - median test, 501
 - one-way ANOVA, 234, 501
 - related samples, 395, 510
 - Durbin test, 517
 - Friedman test, 514
 - Quade test, 510
 - randomized block ANOVA, 395
 - rank ANOVA test, 516
- tests for comparing two locations, 161, 493, 495
 - independent samples, 161, 495
 - Mann-Whitney test, 495
 - two sample t-test, 165, 496
 - Wilcoxon rank-sum test, 495
- paired observations, 176, 493
 - matched pairs test, 493
 - paired t-test, 176, 493
 - sign test, 493
 - Wilcoxon signed ranks test, 493
- tests for comparing variances, 171, 498, 507
 - Bartlett's test, 802
 - Freund-Ansari test, 500
 - F-test, 171, 501
 - k-sample squared ranks test, 507
 - Mood test, 500
 - Siegel-Tukey test, 500
 - squared ranks test, 498
- tests for normality, 531
 - Lilliefors test for normality, 531
 - Shapiro-Wilk test for normality, 532

- tests of single location, 152, 489
 - sign test, 489
 - t-test, 152
 - Wilcoxon signed ranks test, 491
- tests, *see also* hypothesis testing
- theoretical control chart, 211
 - count rate, 213
 - mean, 213
 - number defective, 213
 - proportion, 213
 - range, 216
 - standard deviation, 216
 - variance, 215
- three-sigma control chart, 212
- time series, 21
- time series plot, 21
- tolerance limits, 129
 - distribution free, 129
 - for normal distributions, 129
 - one-sided, 131
 - two-sided, 129
- total bias effect, 676
- traceability of accounting data, 639
- traceability of material, 639
- transformations, 342
- transformations to stabilize variance, 343
- transformed cumulative error model, 775
- treatment means, 235
- t-test, 152
 - two sample, 165, 496
- turning point, 227
- two-way crossed classifications ANOVA, 310
 - computational formulas, 312-313
 - fixed effects model, 313
 - comparison of means, 317-319
 - estimation of means, 317-319
 - confidence intervals, 317-319
 - expected mean squares, 316
 - hypothesis tests, 315
 - mean squares, 313
 - mixed effects model, 324
 - comparison of means, 328
 - estimation of means, 327
 - confidence intervals, 328
 - estimators, 327
 - variance of estimators, 327
 - expected mean squares, 326
 - hypothesis tests, 324-325
 - variance components, 326-327
 - model, 310-311
 - random effects model, 320
 - expected mean squares, 321
 - two-way crossed classifications
 - ANOVA (cont'd)
 - hypothesis tests, 322
 - overall mean, 323
 - confidence interval, 324
 - estimate, 323
 - hypothesis test, 324
 - variance of estimate, 323
 - overall variance, 322
 - variance components, 322
 - sums of squares, 312-313
 - Type I error, 137
 - Type II error, 137
 - types of facilities subject to nuclear
 - material accounting, 623
- unbiased estimator, 98
- uniform distribution, 53, 577
 - parameter estimators, 107
 - pdf, mean, variance, cdf, 53, 54
- union of two sets, 30
- universal set, 30
- unmeasured losses, 643
- upper control limit (UCL), 210
- variable
 - continuous random, 1, 38
 - discrete random, 38
 - independent, 389
- variables measurement, 875
- variables sampling plans, 919
 - required sample size, 922
 - stratified sampling, 921
 - type II error probability, 920
- variance, 42
- variance component estimator, variance of, 699
- variance components, 265, 285, 322, 326, 674, 725
 - negative estimates, 267
- variance estimator
 - combination of, 699
 - variance of, 698
- variance of corrected measurement results, 728
- variance of sample variance, 105
- variance of variance component estimator, 699
- variance of variance estimator, 698
- variance-covariance matrix, 78, 345, 548
- Venn diagram, 31
- verification of accounting data, 637
- verification process, 871
- V-mask, 224

- warning limits, 788
- Weibull distribution, 69
 - pdf, parameters, mean, variance, 68
- weighted average, 7
- weighted average loss, 609
- weighted least squares estimators, 346
- weighted mean, 7
- whiskers, 18
- Wilcoxon rank-sum test, 495
- Wilcoxon signed ranks test, 491, 493
- x-bar, 5
- x-y plot, 20
- Youden square design, 403
- zero acceptance number, 886