23

JUN 05 1985

109.2

~~105.3.1/~~RC/85/05/03

DISTRIBUTION:
~~WM 105.3.1 s/f~~
NMSS r/f
WMGT r/f
DCodell & r/f
MFliegel
JOBunting
MJBell
REBrowning
PDR

MEMORANDUM FOR: Malcolm R. Knapp, Chief, WMGT
Division of Waste Management

FROM: Richard Codell, WMGT
Division of Waste Management

SUBJECT: DEFINITIONS OF "RELIABILITY" AND "CONFIDENCE"

The terms "reliability" and "confidence" appeared several times in our discussions with DOE on Performance Allocations. You asked me to define these terms so that we would be assured that they were being used consistently by both NRC and DOE. The technical or "textbook" definitions of reliability and confidence are presented below:

1. Reliability is... "the probability of a device performing its defined purpose adequately for a specified period of time under the operation condition encountered (Mann et. al. 1974)."

2. Confidence (interval) is "...a statement about an interval within which some unknown constant or parameter must be situated to be in consonance with the available data to some specified extent." (Easterling, 1974) This is usually expressed mathematically as an upper and lower limit determined statistically such that there is at least a 100(1-a)% probability that the interval will contain the true value of the constant or parameter, where "a" is the confidence coefficient (Mann, et. al., 1974).

Discussion

The textbook definition of reliability is rather precise and meaningful in the context of HLW repositories. The definition of confidence is much less so. The dictionary definition of confidence is "Full trust; belief in the reliability of a person or thing." (Random House, 1973) This definition is not adequately conveyed by the textbook definition, which deals only with the statistical properties of a parameter or constant. For example, certain properties of a repository rock such as hydraulic conductivity may be well characterized and we may be confident of their values. We may not be confident of the overall performance of the repository however, because we are uncertain that all important phenomena have been included. The former can be expressed by the textbook definition of confidence, but the latter cannot.

Furthermore, it is important to differentiate between "confidence intervals" and "probability intervals", since the two are often mistakenly used

WM Record File 109.2 WM Project 1
Docket No.
PDR ✓
LPDR
Distribution:
(Return to WM, 623-SS)

OFC	:WMGT	:WMGT	:	:	:
NAME	:RCodell	:MFliegel	:	:	:
DATE	:85/05/				

8507100147 850605
PDR WASTE
WM-1 PDR

interchangeably. A confidence interval is a way of expressing uncertainty about a simple unknown constant. In the analysis of the performance of a repository, for example, the probability of the value of an input parameter to a mathematical model must be assumed in order to perform Monte Carlo simulations. The parameter in question will often be represented as a probability distribution, with the confidence interval representing the probability interval.

A constant is a constant whether we know its value or not, and is not a random variable. It can be quite wrong to probabilistically combine confidence intervals of several constants to obtain a confidence interval on a function of several unknown constants. (Easterling, 1974). Use of confidence intervals as probabilities usually leads to pessimistic confidence intervals of the function. Some mathematicians argue that statistical techniques dealing with confidence intervals should always be used (Mann, et. al, 1974). Others state that, while the above argument may have merit, it is impossible to implement in a practical way for performance assessments (Conover, et.al., 1980).

A few simple examples of the calculation and use of the textbook definition of confidence are presented in the attachment to this memorandum.

Original Signed By

Richard Codell, WMGT
Division of Waste Management

References

1. N.R. Mann, R.E. Schafer, N.D. Singpurwalla, Methods for Statistical Analysis of Reliability and Life Data, John Wiley and Sons, New York, 1974

2. R.E. Easterling, "On the Use of Confidence Intervals as Probability Intervals", USNRC, NR-ASG-001, 1974

3. The Random House College Dictionary, Random House, New York, 1973

4. J. Conover, M. Gillespi, S. Yakowitz, "System Analysis", in Uncertainty Analysis of Post Closure Nuclear Waste Isolation System Performance, Intera Environmental Consultants Inc., Houston Texas, October, 1980

RC

OFC	:WMGT	:WMGT MF	:	:	:	:	:
NAME	:RCodell	:MFliegel	:	:	:	:	:
DATE	:85/06/05	:85/~~05~~/ 06/05	:	:	:	:	:

EXAMPLES OF CONFIDENCE INTERVALS

by Richard B. Codell
Hydrology Section, DWM, NMSS

A few simple examples are presented to illustrate the computation of confidence intervals in the textbook sense.

Confidence on a Mean

The first example deals with the confidence limits of the mean of a sample from a normally distributed population.

The expected value or mean of a population containing N values is:

$$\mu = \left(\sum_{i=1}^{N} x_i\right)/N \tag{1}$$

The mean is itself a constant. A sample mean is the mean calculated from a sample of the population smaller than the entire population:

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n \tag{2}$$

where n is the sample size. The confidence that the sample mean is representative of the mean can be expressed by its confidence interval; providing it is reasonable to assume that the population is normally distributed (Miller and Freund, 1977):

$$\bar{x} - t_{\alpha/2}\, s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}\, s/\sqrt{n} \tag{3}$$

where s is the sample standard deviation

$$s^2 = \left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)/(n-1) \tag{4}$$

$t_{\alpha/2}$ is the Student t distribution statistic for confidence coefficient $\alpha/2$, and n-2 degrees of freedom (a tabulated function). For large samples, s approaches the true standard deviation, σ, and $t_{\alpha/2}$ approaches the normal deviate $z_{\alpha/2}$, also a tabulated function.

Example

We have 12 measurements of a concentration of a radionuclide:

C_i = 15.8, 26.4, 17.3, 11.2, 23.9, 18.7, 13.9, 9.0, 13.2, 22.7, 9.8, 6.2 pci/l.

What is the range within which we have a 95% confidence that the true means will fall?

Solution

The sample mean and sample standard deviation are calculated from Eqs. 2 and 4:

$$\bar{x} = 15.68$$

$$s = 6.339$$

For 95% confidence, $\alpha = 1 - 0.95$, $\alpha/2 = 0.025$. The sample size is 12, so there are (12-2) = 10 degrees of freedom. From a table for the Student t distribution:

$$t_{.025} = 2.228$$

From Eq. 3, therefore:

$$11.60 < \mu < 19.75$$

Confidence on a Standard Deviation

The standard deviation σ of a population is also a constant. The sample standard deviation s is an estimate of σ based on a limited sample of the population. The confidence interval of σ based on s can be expressed:

$$s/(1 + t_{\alpha/2}/\sqrt{2n}) < \sigma < s/(1-t_{\alpha/2}/\sqrt{2n}) \qquad (5)$$

providing it can be assumed that the population is normally distributed.

Example

What is the 95% confidence on the standard deviation of the previous example?

Solution

From the previous example,

$n = 12$, $t_{.025} = 2.228$, $s = 6.339$. Therefore, from Eq. 5:

$$4.357 < \sigma < 11.626 \text{ with 95\% confidence.}$$

Confidence Intervals on A Function of an Independent Variable

Least Squares Estimate

A function of one or several independent variables can often be derived using methods such as least squares. This method derives the best estimate of function y for corresponding values of the independent variables. Confidence intervals on the function can also be derived. If it is assumed for this simple case that y is linearly related to x, and that y is a normally distributed random variable with a constant variance σ^2, the following equations can be stated (all sums from i=1 to i=n):

$$\Sigma y_i = an + b\ \Sigma x_i \tag{6}$$

$$\Sigma x_i y_i = a\ \Sigma x_i + B\ \Sigma x_i^2 \tag{7}$$

where n is the number of data points, and a and b are the coefficients of the regression equation:

$$y = a + bx \tag{8}$$

which expresses the best estimate of y. Equations 6 and 7 are solved simultaneously for a and b.

Confidence Limits on Least Squares Estimate

The confidence limits for estimating a future value of y given x is

$$y = a + bx \pm t_{\alpha/2}\ S_e \sqrt{1 + 1/n + n(x - x)^2/S_{xx}} \tag{9}$$

where $t_{\alpha/2}$ is Student's t statistic with n-2 degrees of freedom for the confidence coefficient $\alpha/2$,

S_e is the standard error of estimate,

$$S_e^2 = (S_{xx}S_{yy} - (S_{xy})^2)/(n(n-2)S_{xx}) \tag{10}$$

and,

$$S_{xx} = n\ \Sigma x_i^2 - (\Sigma x_i)^2 \tag{11}$$

$$S_{yy} = n\ \Sigma y_i^2 - (\Sigma y_i)^2 \tag{12}$$

$$S_{xy} = n\ \Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i) \tag{13}$$

(all sums from i=1 to i=n)

<u>Example</u>

A simple example of the use of the least square method and the prediction of confidence intervals is presented below (Miller and Freund, 1977):

Consider the tabular data of independent variable x and dependent variable y.

x	y
20	0.18
60	0.37
100	0.35
140	0.78
180	0.56
220	0.75
260	1.18
300	1.36
340	1.17
380	1.65

Calculating the summed terms and substituting into Eqs (6) and (7) gives:

$$8.35 = 10\ a + 2000b$$
$$175.4 = 2000a + 532000b$$

which can be simply solved for a and b and substituted into Eq. 9:

$$y = 0.069 + 0.0038x$$

The confidence limits are calculated from Eq. 13. The t statistic is taken from a table for $\alpha/2 = 0.025$ and $(10-2) = 8$ degrees of freedom.

$$t_{.025} = 2.306$$

The data, least squares fit and 95% confidence bands are shown on Fig. 1. The confidence bands are narrowest at the center of the graph, corresponding to the mean value of x, and widest at the ends.

Reference

I. Miller and J. Freund, Probability and Statistics for Engineers, Prentice Hall, Englewood Cliffs, NJ, 1977

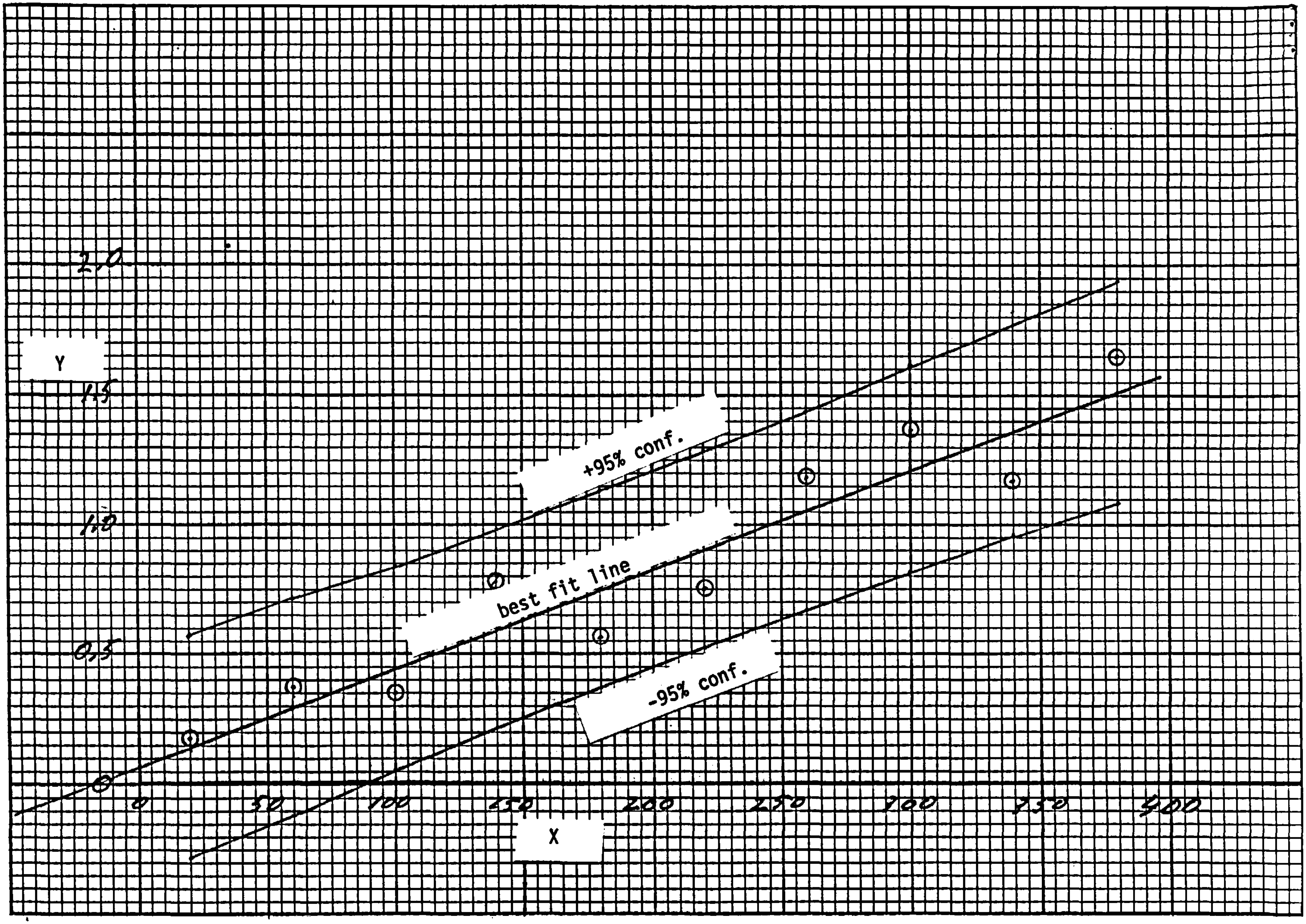

Figure 1 - Linear Regression Example