Section 3: Frequentist Statistical Inference

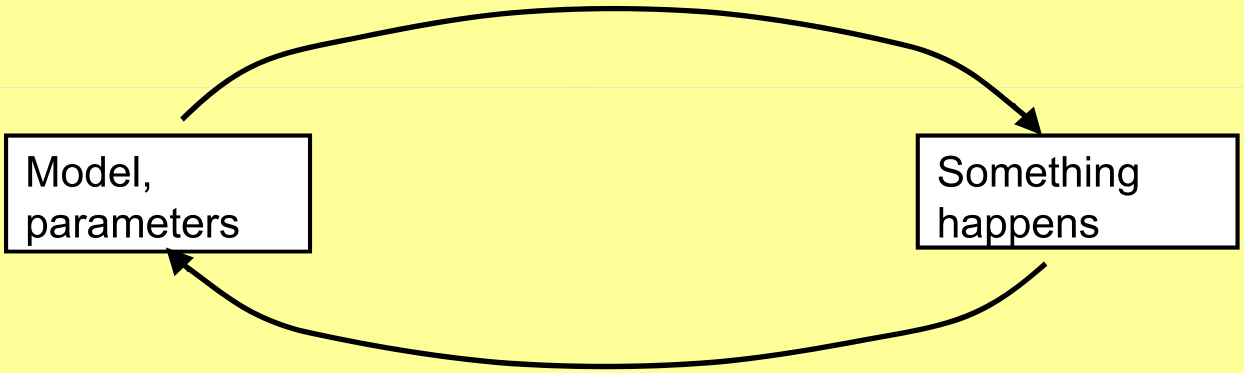
- *Purpose*
  - *Students will learn about maximum likelihood estimators, confidence intervals, and about methods of model validation*
- *Objectives*
  - *Students will learn*
    - *Definition of maximum likelihood estimator (MLE) and confidence interval*
    - *Application of these estimators to Poisson, binomial, and exponential data*
    - *Graphical tools for model validation*
    - *Intro to hypothesis-testing for model validation, with example applications*

Difference between Probability and Statistical Inference

Probability

Find Pr(something happens)

Events are stochastic



Statistical Inference

Decide on plausible model
Estimate its parameters

Conclusions contain uncertainty

Frequentist Statistical Inference

- *Estimation*
  - *Point Estimates*
  - *Interval Estimates (Confidence Intervals)*
- *Model Validation*
  - *Graphical Methods*
  - *Tests of Hypotheses*

LOSP Example Data

- *In the LOSP example, suppose that we have collected data related to this type of event*
- *The rate of experiencing LOSP*
  - *1 initiating event in 9.2 operating years*
- *The probability of not starting a diesel generator (DG)*
  - *1 failure to start in 75 demands*
- *The rate of a DG not operating*
  - *0 failures to run in 146 running hours*

Tips for Solving Many Statistics Problems

- Answer the following questions, in order:
  - What are the data?
  - Which part(s) of data considered to be observations of a random variable?
  - Which model (distribution) generated the observed data?
    - State which parameter(s) of distribution are known/which are unknown
  - For Bayesian analysis, what is prior distribution of unknown parameter(s)?
  - What is to be found?
    - Point estimate, confidence interval, Bayes posterior distribution, Bayes credible interval, ...
    - Find the formula for it in HOPE, course slides, or notes
- Replace symbols in the formula by known quantities from the data. Look up any needed values from tables, and write out the answer!

Point Estimation

- *Point* ***estimator***
  - *A function of random data that estimates unknown parameter*
- *Point* ***estimate***
  - *Value of estimator for actual observed data*
- *That is*
  - *Estimator is a random variable (upper case, Λ)*
  - *Estimate is a number, the value the random variable takes (lower case, λ)*

HOPE

Pages B-1 through B-4

Likelihood Function

- *For discrete random variable*
  - ***Likelihood** = Pr(data), considered as a function of the unknown parameter(s)*
- *For continuous random variable*
  - ***Likelihood** = density of the data, considered as a function of the unknown parameter(s)*
- *In other words, the **likelihood** is the pdf written as a function of the unknown parameter(s)*

Likelihood — Discrete Example

- *Recall for Poisson and binomial distributions (both discrete)*

$$\Pr(X = x) = e^{-\lambda t} (\lambda t)^x / x!$$

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

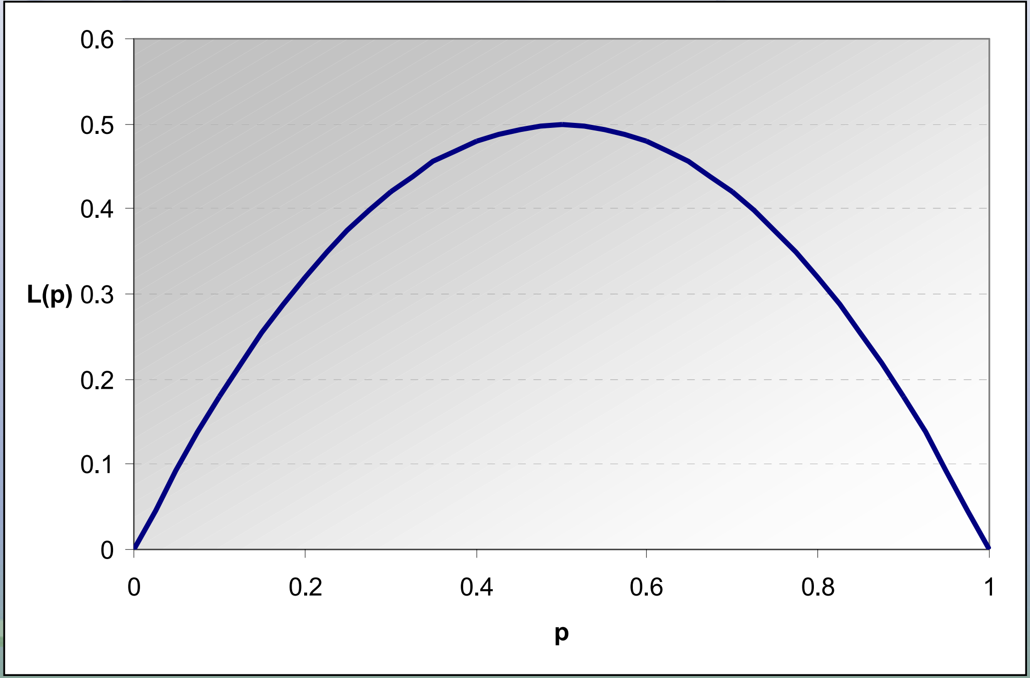
- *Binomial likelihood example*
  - *One failure out of two trials, x = 1, n = 2*

$$L(p) = \Pr(X = 1 \mid p) = \binom{2}{1} p^1 (1-p)^{2-1}$$

$$= \frac{2!}{1!(2-1)!} p(1-p) = 2p(1-p)$$

Likelihood — Discrete Example (cont.)

- What does this binomial likelihood look like?
  - One failure out of two trials, $x = 1$, $n = 2$
  - $L(p) = 2p(1 – p) = 2p - 2p^2$
- The question then is
  - What value of p "maximizes" L(p)?
  - In other words, what value of p most likely yields data of one failure out two trials



Likelihood — Continuous Example

- *Recall for exponential distribution (continuous)*

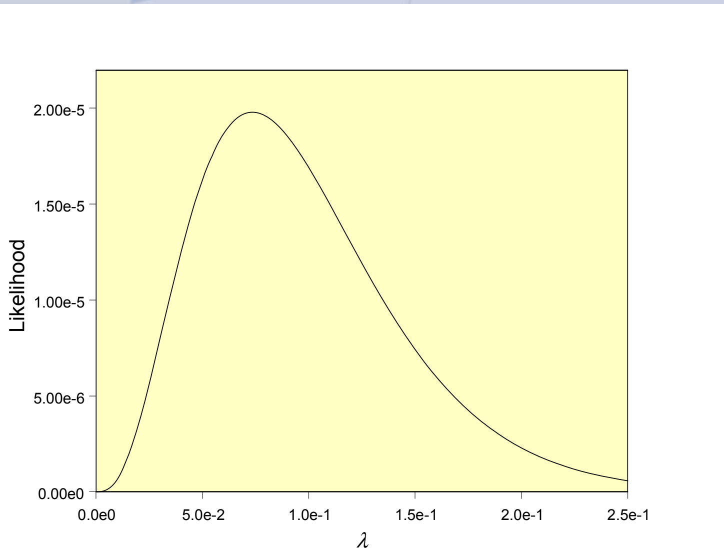
$$f(t) = \lambda e^{-\lambda t}$$

- *Observe n **independent** event times, each with same exponential distribution.*
  - *The observations are $t_1, t_2, \ldots, t_n$*
  - *The likelihood is*

$$L(\lambda) = \prod f(t_i \mid \lambda)$$

$$= \prod \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \Sigma t_i}$$

  - *Example with n=3, Σt_i=40.8*



Maximum Likelihood Estimate (MLE)

- *Given the data, the MLE is the parameter value that maximizes the likelihood function*
  - *This is what we saw on the likelihood example*
- *Formulas for common cases*
  - *If X ~* **Poisson**(λt)*, and x is observed, the MLE is*

    $$\hat{\lambda} = x / t$$

  - *If X ~* **binomial**(n, p)*, and x is observed, the MLE is*

    $$\hat{p} = x / n$$

  - *If* T_i *~* **exponential**(λ)*, independently distributed,* $i = 1,\ldots,n$*, and* $t_1,\ldots, t_n$ *are observed, the MLE is*

    $$\hat{\lambda} = n / \Sigma t_i$$

MLE (cont.)

- *Maximum likelihood estimates for LOSP example*
  - *If X ~* **Poisson**(*λt*)*, and x=1 event (in 9.2 years), the MLE is*

    $$\hat{\lambda} = 1/9.2\,years = 0.11\,per\,year$$

  - *If X ~* **binomial**(*n, p*)*, and x=1 failure (in 75 demands), MLE is*

    $$\hat{p} = 1/75 = 0.013$$

  - *If X ~* **Poisson**(*λt*)*, and x=0 failures (in 146 hours), the MLE is*

    $$\hat{\lambda} = 0/146\,hours = 0\,per\,hour$$

Moment Estimates

- *Estimate moments by the corresponding sample moments*
  - *Sample mean*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  - *Sample variance*

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

- *In this course, these are useful for estimating moments of distributions produced by Monte Carlo simulations, when n is large*

Confidence Intervals

- *A 100(1 $-\alpha$)%* ***confidence interval*** *has the form (L, U), where L and U are functions of the data*
- *The interval satisfies:*
  - *Pr(L ≤ parameter ≤ U) ≈ 100(1 $-\alpha$)%*
  - *We would like exact equality, but with discrete data we have to settle for*

$$Pr(L \le \text{parameter} \le U) \ge 100(1-\alpha)\%$$

- *IMPORTANT*
  - *In this equation, the parameter is fixed, and L and U are considered random.*

HOPE

Pages 6-6, 6-7, 6-32, 6-33, 6-51, B-5, B-6

Formulas for 90% confidence intervals (with analogous formulas for other confidence levels)

- *If X ~ Poisson(λt),*

  $(L, U) = (\chi^2_{0.05}(2X)/(2t), \chi^2_{0.95}(2X + 2)/(2t))$

- *If X ~ binomial(n, p),*

  $(L, U) = (beta_{0.05}(X, n - X + 1), beta_{0.95}(X + 1, n - X)$

- *If T_i ~ exponential(λ), independently distributed, i = 1,…,n, and $T = \Sigma T_i$,*

  $(L, U) = (\chi^2_{0.05}(2n)/(2T), \chi^2_{0.95}(2n)/(2T))$

- *One-sided intervals are also possible (L, ∞) or (–∞, U).*

Summary So Far of Estimates for LOSP Example

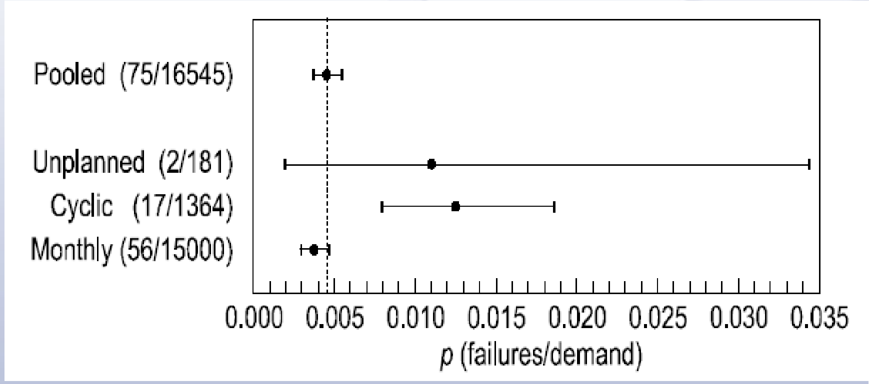
Parameter	Data	MLE	90% Conf. Int.
λ_{LOSP}	*1 event in 9.2 op. yrs.*	$1.1E\text{-}1\ yr^{-1}$	$(5.6E\text{-}3,\ 5.2E\text{-}1)\ yr^{-1}$
p_{FTS}	*1 failure in 75 demands*	$1.3E\text{-}2$	$(6.8E\text{-}4,\ 6.2E\text{-}2)$
λ_{FTR}	*0 failures in 146 hrs*	$0\ hr^{-1}$	$(0\ ,\ 2.1E\text{-}2)\ hr^{-1}$

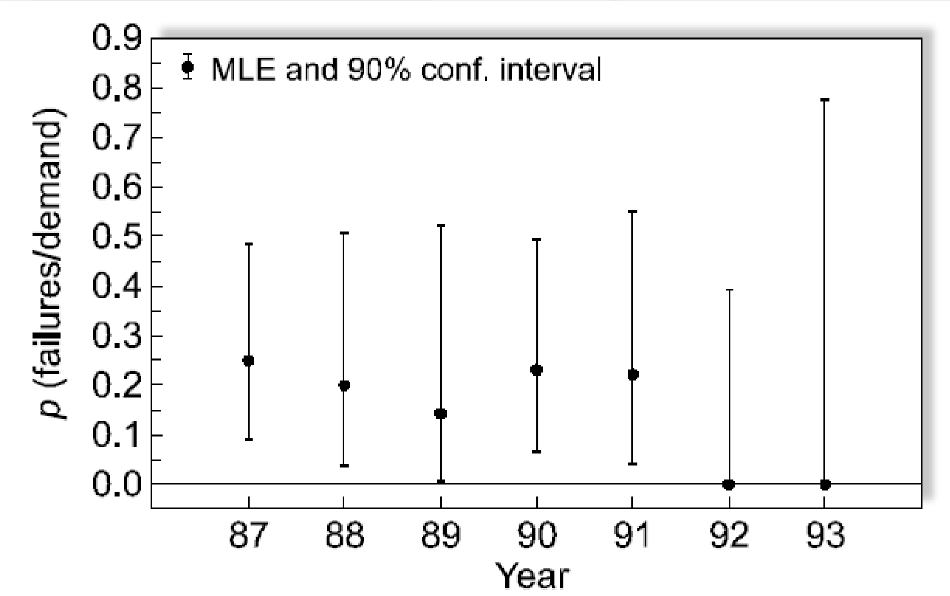
MODEL VALIDATION

- *It is important to check the model assumptions, by seeing if the data and model are consistent — you can make big errors if you forget to check model assumptions*
- *Graphical methods provide insights about possible violations of assumptions*
- *Hypothesis tests quantify the strength of the evidence against the assumptions*

Two simple graphs are useful

- *Side-by-side confidence intervals*

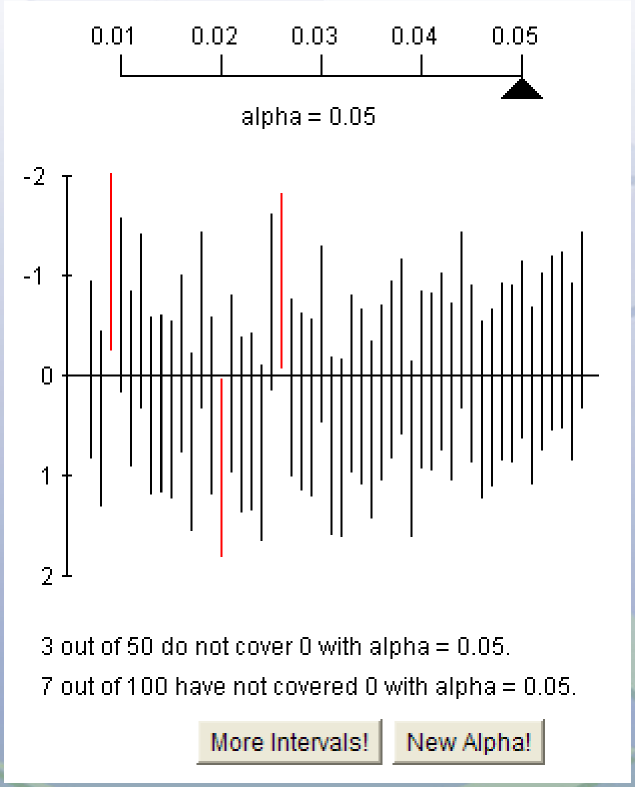




Pages 6-21, 6-25, 6-48

Two simple graphs are useful (cont.)

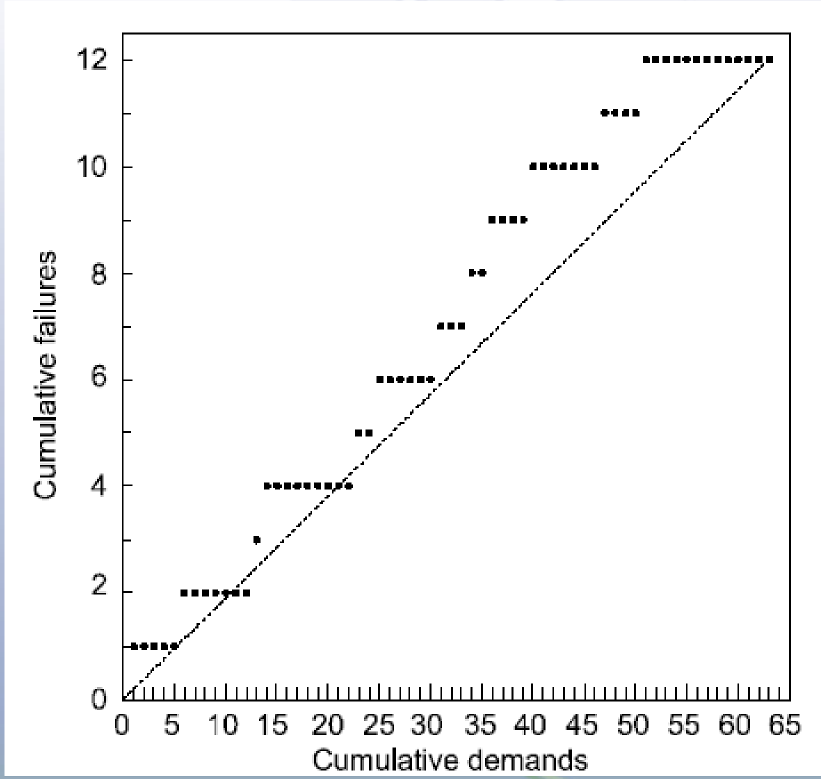
- *Side-by-side confidence intervals*

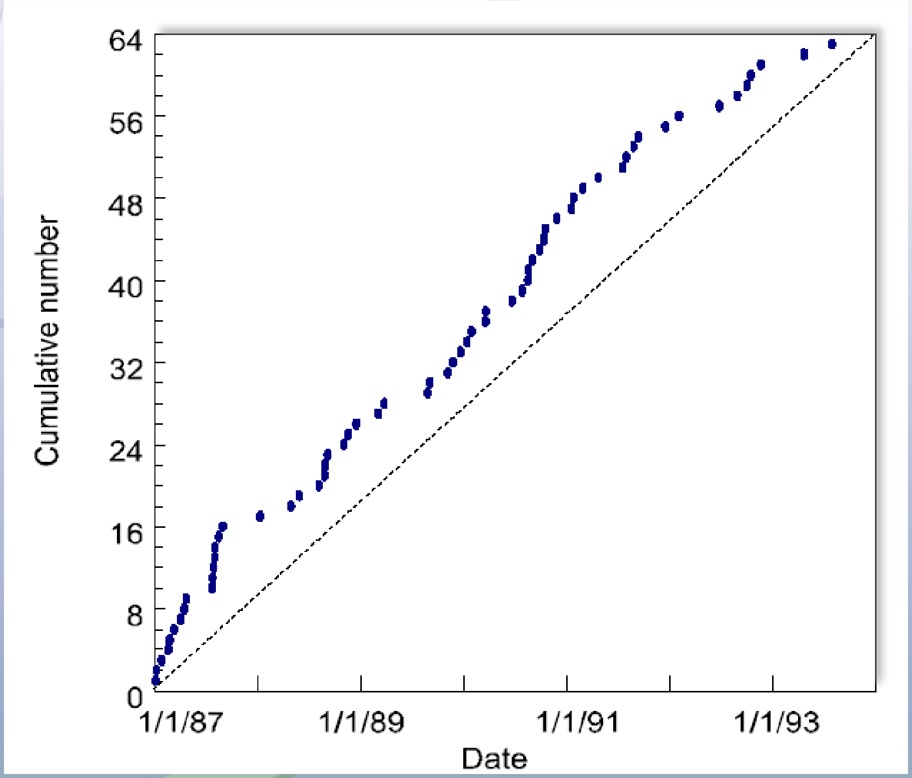


http://www.stat.sc.edu/%7Ewest/javahtml/ConfidenceInterval.html

Two simple graphs are useful (cont.)

- *Cumulative plots*





HOPE

Pages 6-25, 6-48

Test of a Hypothesis – Framework

- *Define two possibilities, or hypotheses*
  - *H_0 the null hypothesis*
    - *A simple assumption that is used unless the data give good reason not to believe hypothesis*
  - *H_1 the alternative hypothesis*
    - *A more complex assumption that will be used only if the null hypothesis is rejected*
- *But, two kinds of error are possible*
  - *Type I error — reject H_0 when H_0 is true*
  - *Type II error — accept H_0 when H_0 is false*

HOPE

Pages B-6 through B-8

Test of a Hypothesis – Decision Rule

- *The "decision rule"*
  - *Construct test statistic (Y), which is a function of the data*
- *Decide to limit Pr(Type I error)*
  - *Set α = desired Pr(reject H_0 | H_0 true)*
  - *A typical value of α is 0.05 (other values are possible)*
  - *The more important the decision, in general the smaller one would want to set α*
    - *This probability of false alarm is the chance you make wrong inference when original hypothesis was correct*

Test of a Hypothesis – Decision Rule

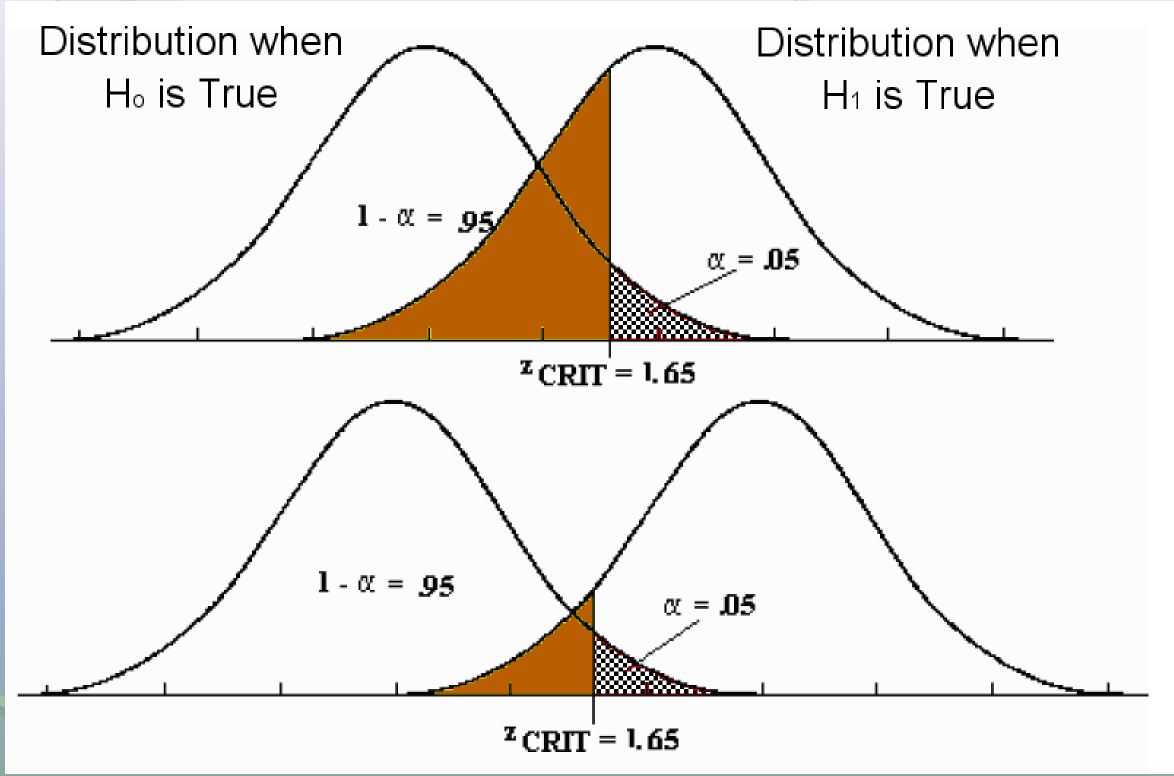
- Construct some "critical region", with

$$Pr(Y \text{ in critical region} \mid H_0 \text{ true}) = \alpha.$$

  – If α is small, we do not expect to see Y fall in critical region, unless H_0 is false.
- Collect data, calculate the value of Y
  – If calculated value of Y is in critical region, **reject** H_0 in favor of H_1
  – Otherwise, give H_0 benefit of doubt, "**accept**" H_0 (do not reject the null hypothesis)

Test of a Hypothesis – Graphical Interpretation

- *What we are really "testing" are two sampling distributions, one for H_0 and one for H_1*
- *When these two distributions change, then decision may change*
  - *Example when the mean value varies (two data sets)*

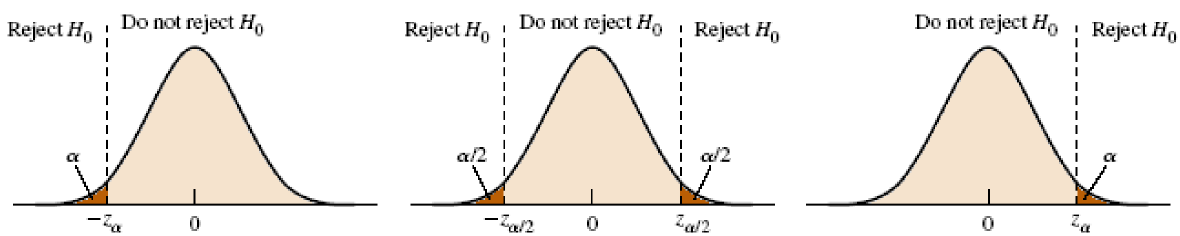


Test of a Hypothesis – Summary

- *From the "Pocket Dictionary of Statistics"*

hypothesis and an **alternative hypothesis.** A null hypothesis is usually tested and either rejected in favor of an alternative hypothesis or not rejected, in which case the alternative hypothesis cannot be sustained.

1. State a null hypothesis (H_0) based on the specific question or phenomenon to be investigated.
2. State an alternative hypothesis. This may be one-sided or two-sided depending on the problem being investigated as defined in the null hypothesis.
3. Specify the **level of significance** (α). This is commonly taken as 0.05 and represents the maximum acceptable **probability** of incorrectly rejecting the null hypothesis.
4. Determine an appropriate **sampling distribution** of the **sample statistic** of interest. Select a one-tailed or two-tailed test, depending on the alternative hypothesis.
5. Evaluate the **standard error** or, more generally, an **estimate** of the standard error of the sample statistic; the formula for the standard error depends on the sample statistic in question.
6. Compute the true value of the **test statistic** and locate its value on the sampling distribution.
7. Reject or do not reject H_0, depending on whether or not the sample statistic is located on the sampling distribution at or beyond the value of the test statistic at a given α.



Test of a Hypothesis – Another Way of Stating Conclusions

- *The **p-value***
  - *Value of α at which H_0 would just **barely** be rejected*
    - *In other words, "shift" α left or right until H_0 is not accepted*
- *It is common to collect data, calculate y, and then report the p-value*
  - *This approach measures strength of evidence against H_0*
  - *A small p-value corresponds to strong evidence against H_0*
- *A p-value < 0.05 corresponds to "statistical significance"*

Commonly Used Tests

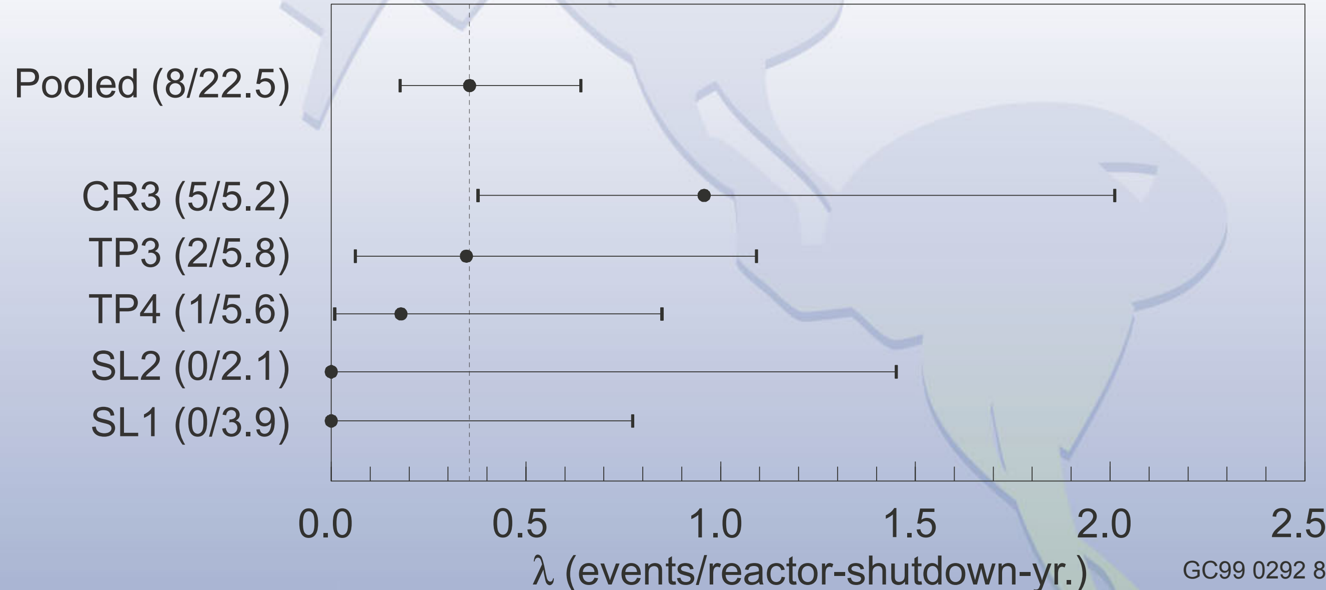
- To test H_0: $\lambda = \lambda_0$
  - Reject H_0 if confidence interval for λ does not contain λ_0
- The "chi-squared test" is one test of more complicated hypotheses.
- Many other tests exist for various situations

HOPE

Pages 6-22 and 6-23

Chi-Squared Test — Example

- *LOSP was considered for 5 plants while in shutdown*



- *We want to test the hypothesis*

  H_0: λ is same at all five plants

  H_1: λ is not same at all five plants

Chi-Squared Test — Example (cont.)

- *The data, and start of calculations for test*

Plant	Events, x_i	Plant shut-down yrs, t_i	Expected count, $e_i = \hat{\lambda} t_i$
CR3	5	5.224	1.857
SL1	0	3.871	1.376
SL2	0	2.064	0.734
TP3	2	5.763	2.048
TP4	1	5.586	1.985
Totals	8	22.508	8.000

$$\hat{\lambda} = \Sigma x_i / \Sigma t_i = 0.355$$

HOPE

Pages 6-22 and 6-23

Chi-Squared Test — Example (cont.)

- e_i is expected count assuming H_0 true
- Compare observed, x_i, with expected, e_i
- Combine the results for each cell by

$$X^2 = \sum \frac{(x_i - e_i)^2}{e_i}$$

- If H_0 true, distribution of X^2 is approximately chi-squared, if e_i values are not too small
  - Degrees of freedom = no. of cells -1, i.e. 4 in example
  - For "not too small", see HOPE, p. 6-24
- If H_0 false, X^2 tends to be larger than chi-squared random variable. So reject H_0 if X^2 in in tail of chi-squared distr.

Chi-Squared Test — Example (cont.)

- *In this example, we can calculate $X^2 = 7.92$*
- *Close to 90th percentile of $\chi^2(4)$ distribution*
- *Conclusion: evidence is borderline, but not quite strong enough to justify rejecting H_0.*
- *This conclusion was also suggested by the picture*
  - *But interpretation of the picture is somewhat subjective*
  - *Hypothesis test objectively quantifies the strength of the evidence against H_0*

Hypothesis Testing Example – Maintenance Rule Performance Criteria

- *Component has assumed probability of failure (PRA point estimate)*
- *Component tests or demands can reveal failure*
- *How many failures are too many for assumed probability of failure to be accepted?*
- *Null hypothesis – H_o: $p = p_o$*
- *Alternative hypothesis – H_1: $p > p_o$*
- *Choose $\alpha = 0.05$*

Hypothesis Testing Example – Maintenance Rule Performance Criteria

- *Assume 24 tests will be performed over next evaluation period*
- *Assume number of failures follows binomial distribution with n =24 and p = p_o (under H_o). Assume p_o = 0.06.*
- *Need to find x_{crit}, such that Pr(rejecting H_o| H_o true) = 0.05.*
- *That is, $Pr(X > x_{crit}|n=24, p_o = 0.06) = 0.05$*
- *Easier to use $Pr(X \leq x_{crit}|n=24, p_o = 0.06) = 0.95$*
- *Because X is discrete, don't strive for exact equality in the above*

Hypothesis Testing Example – Maintenance Rule Performance Criteria

- *Using binomial distribution, find $Pr(X \leq 3 | n=24, p_o = 0.06) = 0.947$, so performance criterion is set at 3 or fewer failures in 24 tests.*
- *What is Type II error probability?*
  - *Depends on what p actually is*
  - *Assume $p_{act} = 3\ p_o = 0.18$*
  - *$Pr(\text{Type II error}) = Pr(\text{accepting } H_o | H_1 \text{ true}) = Pr(X \leq 3 | n=24, p_{act} = 0.18) = 0.35$*

Summary of Frequentist Estimates for LOSP Example

Parameter	Point Est. (MLE)	90% Interval
λ_{LOSP}	*1.1E-1 yr* $^{-1}$	*(5.6E-3, 5.2E-1) yr* $^{-1}$
p_{FTS}	*1.3E-2*	*(6.8E-4, 6.2E-2)*
λ_{FTR}	*0 hr* $^{-1}$	*(0 , 2.1E-2) hr* $^{-1}$