Sample Size Calculations

Paul Mathews
Mathews Malnar and Bailey, Inc.
466 West Jackson Street, OH 44077
Phone: 440-350-0911
Website: www.mmbstatistical.com
E-mail: paul@mmbstatistical.com

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

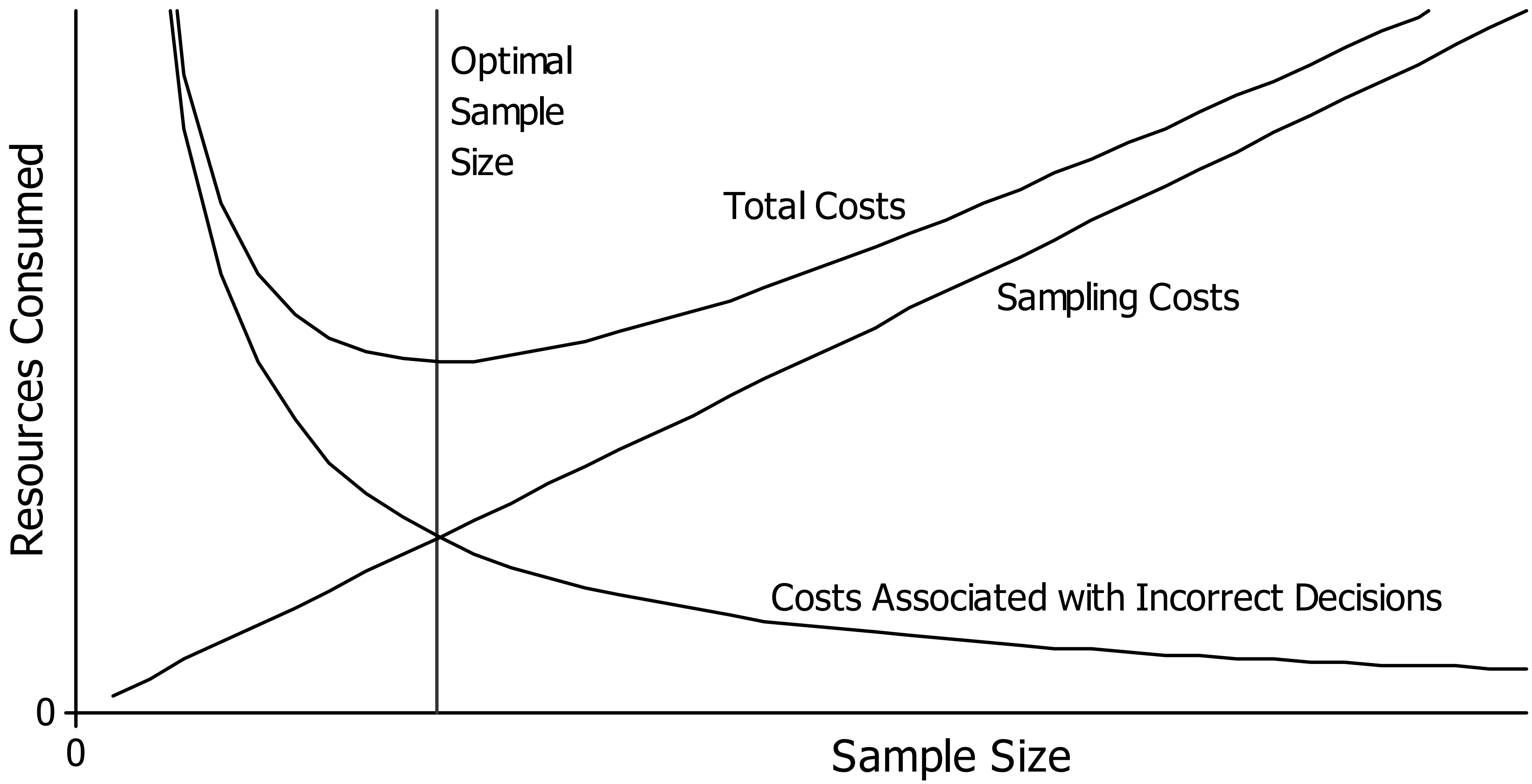
Getting to Know You

- Who are you?
  - Name
  - Title
  - Company
- Do you perform sample size and power calculations for your organization or are you just here for the CEUs?
- Are sample sizes set objectively in your organization or are they based on arbitrary or historical choices?
- Do you use sample size or power software?
- Do you use published standards?

Seminar Outline

1. **Review of Fundamental Concepts**
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Motivation



Optimal
Sample
Size
Total Costs
Sampling Costs
Costs Associated with Incorrect Decisions
Resources Consumed
0
0
Sample Size

Definition

An *experiment* is any activity that involves data collection, analysis, and interpretation for the purpose of making decisions about how to manage a process.

If an experiment is worth doing, it should be done with the right sample size.

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of." - Sir Ronald Fisher

Point of View

- We take the point of view of a statistical consultant who is expected to provide technical support for sample size and power calculations but who has no experience with or knowledge of the process.
- The statistical consultant is dependent on the researcher for:
  - Information about the process
  - The limitations and goals of the experiment
  - Executing the experiment
  - Reporting deviations from the experiment plan
  - Investigating unusual observations
  - Recommending first principles to guide the analysis of the data
  - Interpreting the results for practical significance

Software

- Piface - *www.stat.uiowa.edu/~rlenth/Power/*
- PASS - *www.ncss.com*
- MINITAB - *www.minitab.com*
- R - *www.r-project.org*

- An important trick: Most sample size software does calculations for hypothesis tests but not for confidence intervals. To do the sample size calculation for a confidence interval set the power to 50% in the sample size calculation for a hypothesis test.

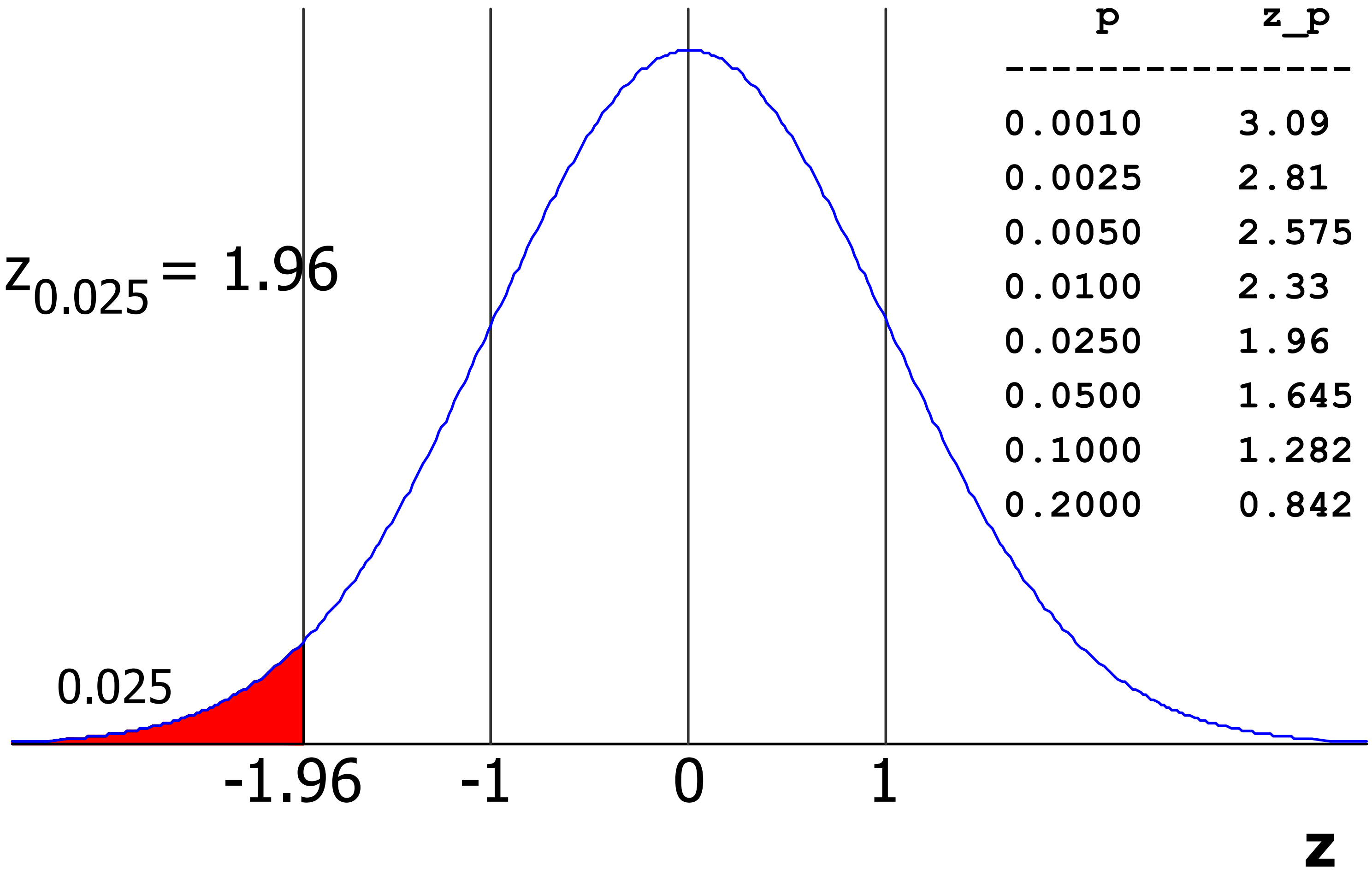
Approximate Methods

- Factors of two are only important in matters of salary (at least in physics).
- Power and sample size calculations are done when there are significant uncertainties in inputs to the calculations, so approximate calculation methods may be tolerated.
  - Ignore the continuity correction for discrete random variables
  - Use large sample approximations with standard deviations determined by the delta method
  - Everything looks normal if the sample size is big enough
- Sources of uncertainties:
  - Value of the standard deviation
  - Value of the confidence interval half-width and confidence level
  - Value of hypothesis test effect size and power
  - Knowledge of the process
  - Likelihood that the experiment will go as planned
  - Validity of the analysis method
  - Assumption violations

Probability Distributions

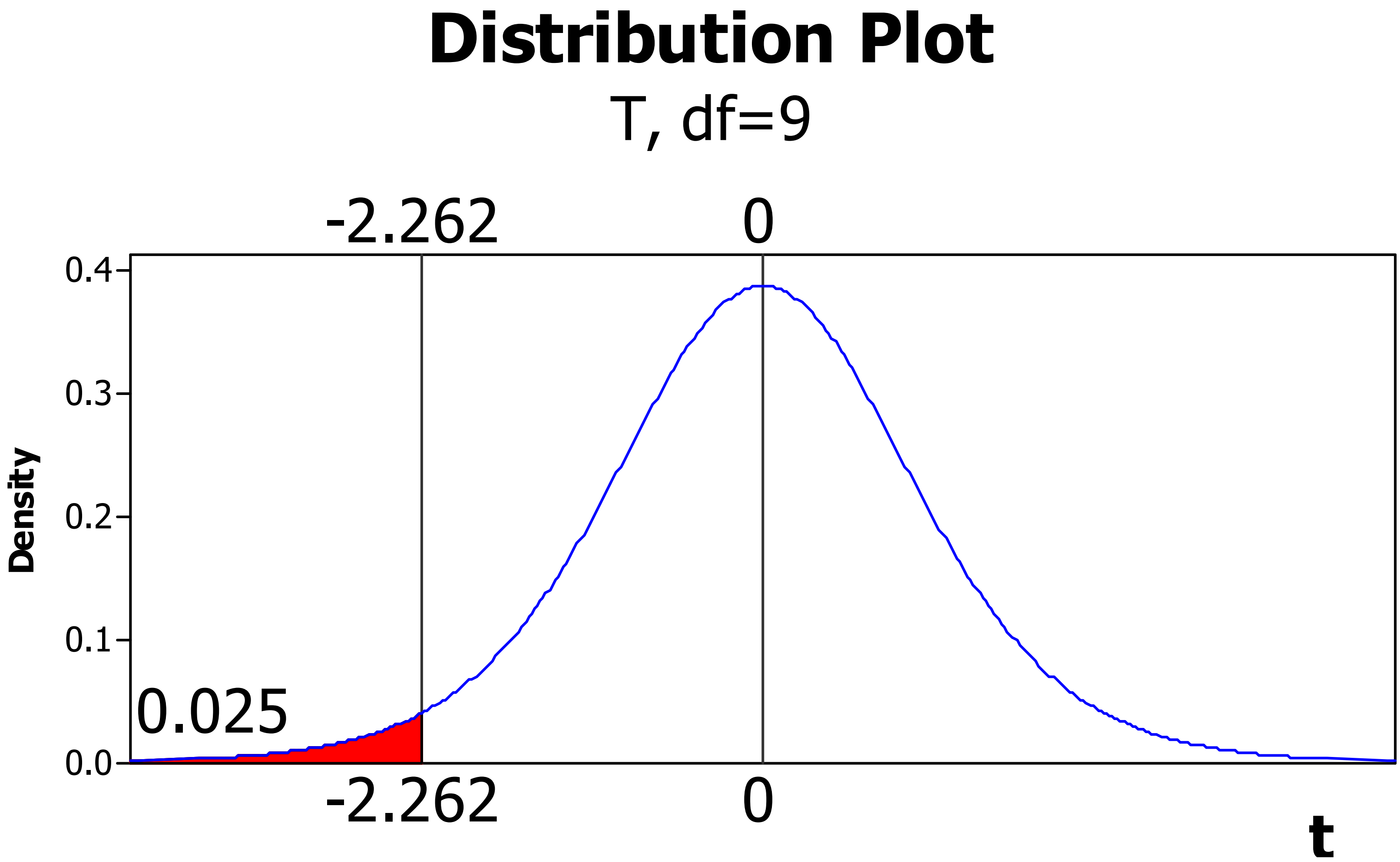
- Normal
- Student's *t*
- Chi-square (χ^2)
- *F*

Normal Distribution

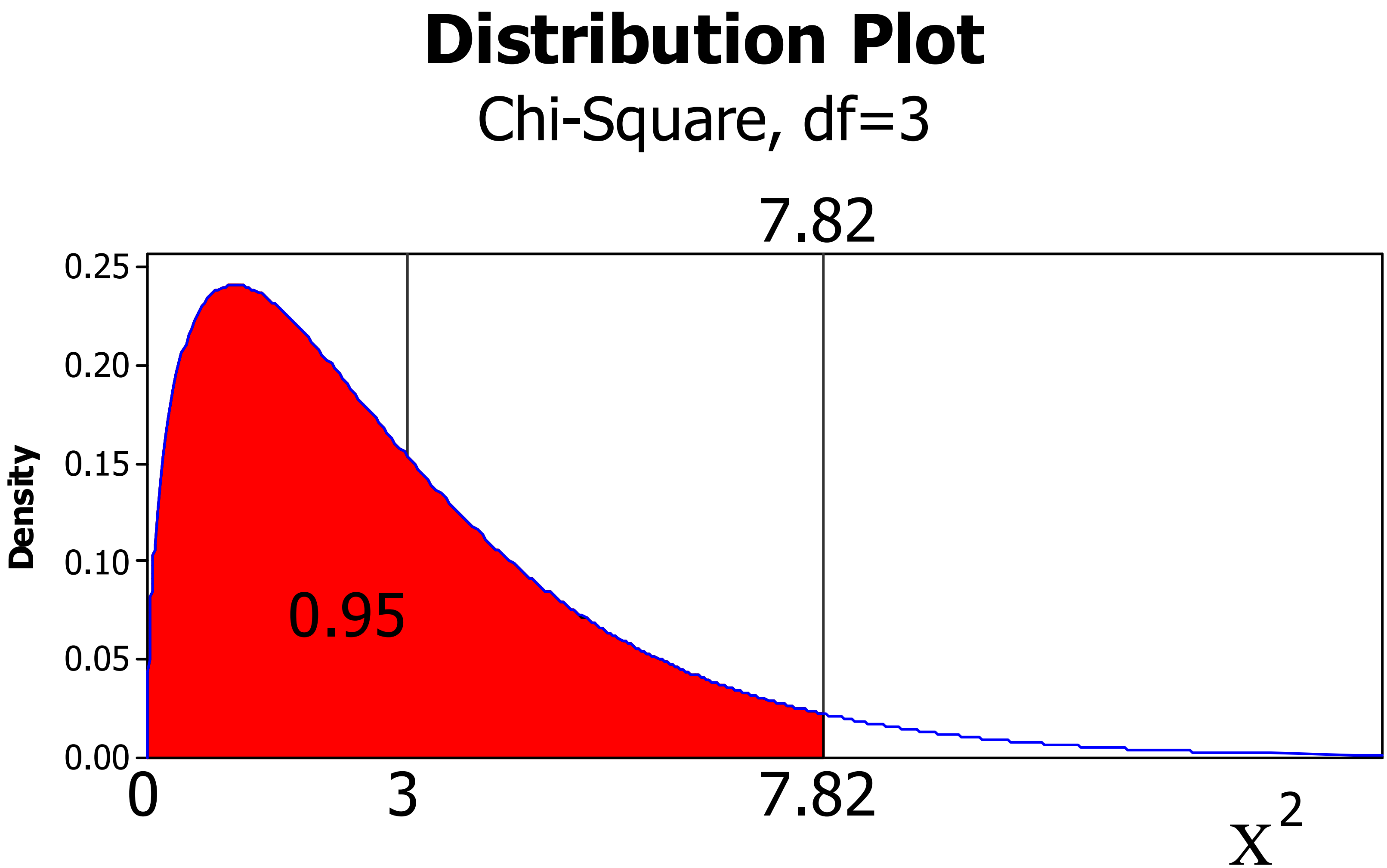


p	z_p
0.0010	3.09
0.0025	2.81
0.0050	2.575
0.0100	2.33
0.0250	1.96
0.0500	1.645
0.1000	1.282
0.2000	0.842

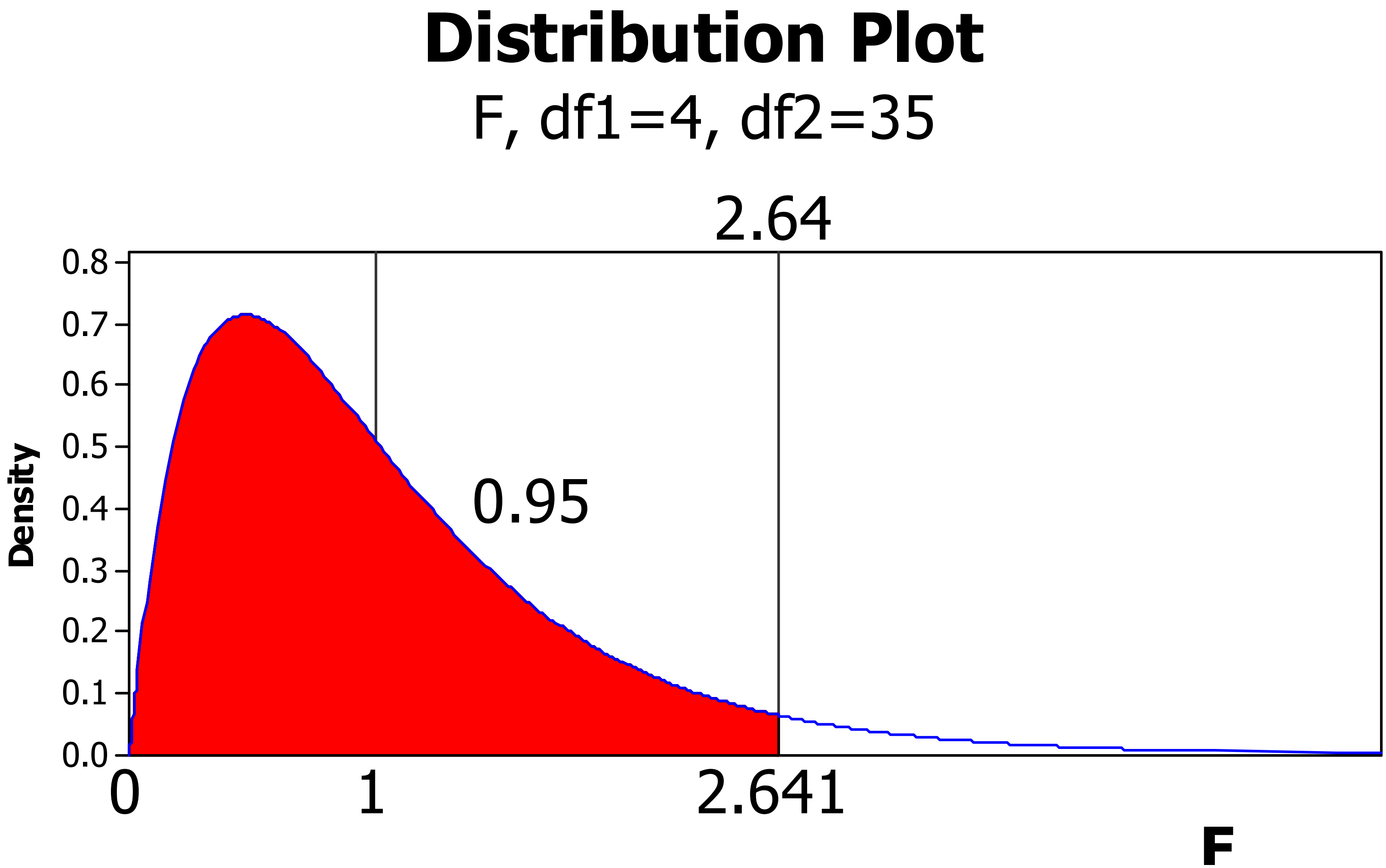
Student's *t* Distribution



Chi-Square (χ^2) Distribution



F Distribution



Delta Method

- Probability distributions often have undesirable properties such as skewness or heteroscedasticity.
- For many of these distributions, a mathematical transformation of the original random variable results in a distribution that is better behaved.
- Transforming a badly behaved distribution into a better behaved one allows simpler and better known analysis methods to be used.
- The *delta method* is used to estimate the standard deviation of a transformed distribution from the original distribution.
- It’s not necessary to understand how to apply the delta method; however, it’s very helpful to make use of its results.

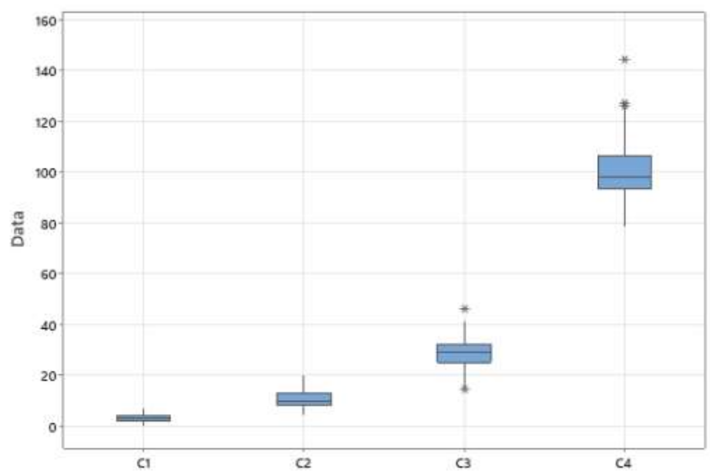
Delta Method

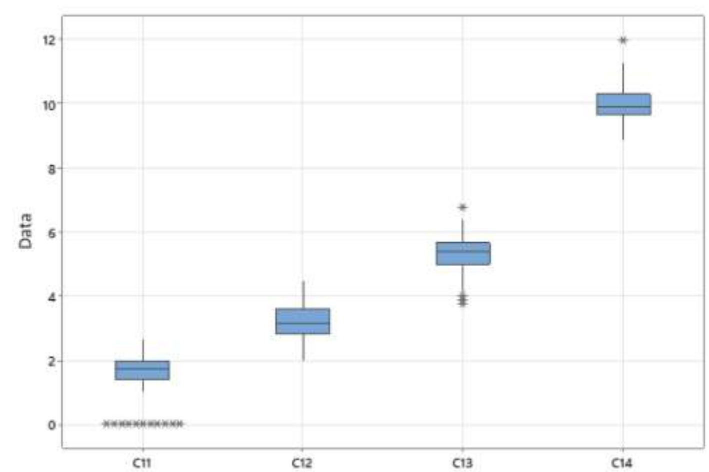
Example: When the distribution of sample counts x is Poisson with mean $\mu_x = \lambda$ and standard deviation $\sigma_x = \sqrt{\lambda}$, the sampling distribution of $\sqrt{x}$ is approximately normal with mean

$$\mu_{\sqrt{x}} = \sqrt{\lambda}$$

and standard deviation

$$\hat{\sigma}_{\sqrt{x}} = \sigma_x \left(\frac{d}{dx} (\sqrt{x}) \right|_{x=\lambda} = \sqrt{\lambda} \left(\frac{1}{2\sqrt{\lambda}} \right) = \frac{1}{2}.$$





Delta Method

Example: An approximate 95% confidence interval for λ is

$$P\left(\sqrt{x} - z_{0.025}\hat{\sigma}_{\sqrt{x}} < \sqrt{\lambda} < \sqrt{x} + z_{0.025}\hat{\sigma}_{\sqrt{x}}\right) = 0.95$$

$$P\left(\sqrt{x} - 1 < \sqrt{\lambda} < \sqrt{x} + 1\right) = 0.95$$

$$P\left((\sqrt{x} - 1)^2 < \lambda < (\sqrt{x} + 1)^2\right) = 0.95$$

Delta Method

Example: Determine the 95% confidence interval for the Poisson mean if $x = 25$ counts are observed.

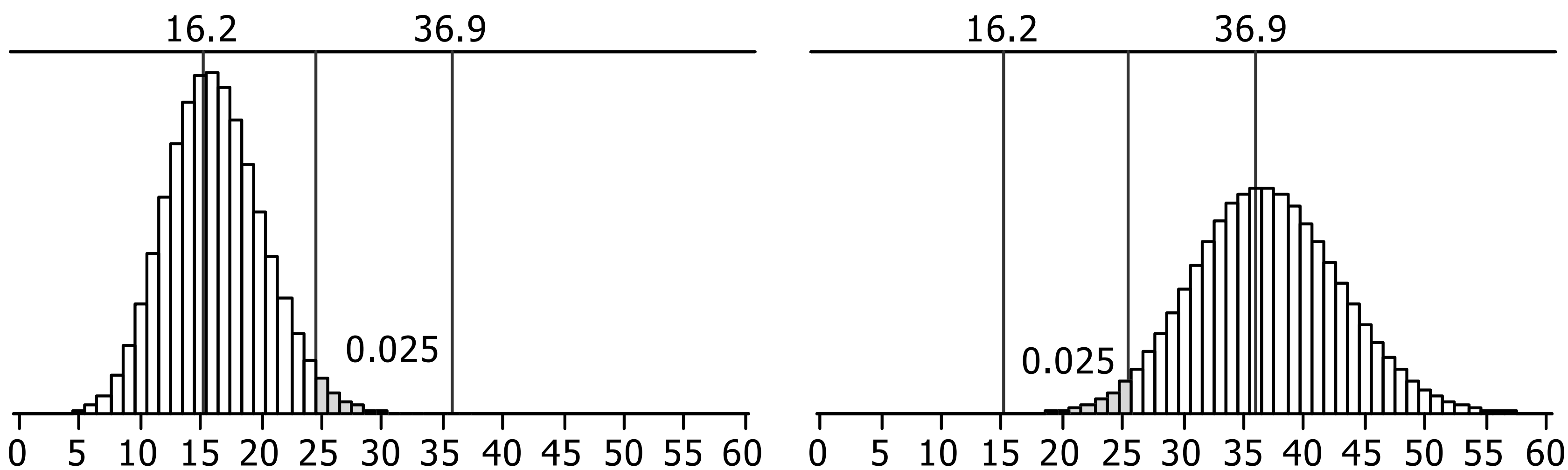
Solution: By the delta method:

$$P\left(\left(\sqrt{25} - 1\right)^2 < \lambda < \left(\sqrt{25} + 1\right)^2\right) = 0.95$$

$$P(16 < \lambda < 36) = 0.95$$

The exact confidence interval is:

$$P(16.2 < \lambda < 36.9) = 0.95$$



Sample Size Calculations

To calculate the sample size, we must know the:

- Purpose of the experiment
- Type of data to be collected
- Parameter to be studied
- Experiment design
- Intended statistical analysis and decision criteria
  - Confidence interval half-width, confidence level
  - Effect size and power, significance level
- Population standard deviation

Sample Size Myths

- There are many sample size myths that are not true
- Assuming variable data, it's common to see statements like "It is generally accepted that the sample size $n = 30$ is sufficient."
  - This statement is false ...
  - Because the sample size must be matched to an intended analysis method ...
  - So there are different sample sizes required for studying the mean, standard deviation, proportion defective relative to a specification limit, process capability, and distribution shape.

Seminar Outline

1. Review of Fundamental Concepts
2. **Means**
    a. **One Mean**
    b. **Two Means**
    c. **Many Means**
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

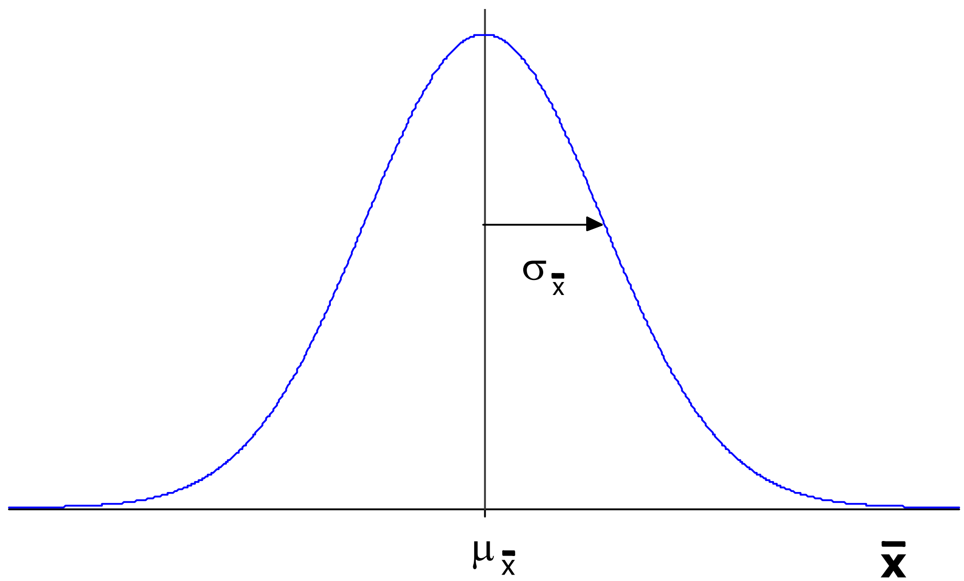
Confidence Interval for the Mean

The Central Limit Theorem says that the distribution of sample means ($\bar{x}$) is normal with mean $\mu_{\bar{x}} = \mu_x$ and standard deviation $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$. This result can be used to construct probability statements about the range of $\bar{x}$ values such as

$$\Phi(\mu_x - \delta < \bar{x} < \mu_x + \delta) = 1 - \alpha$$

where the interval half-width δ is

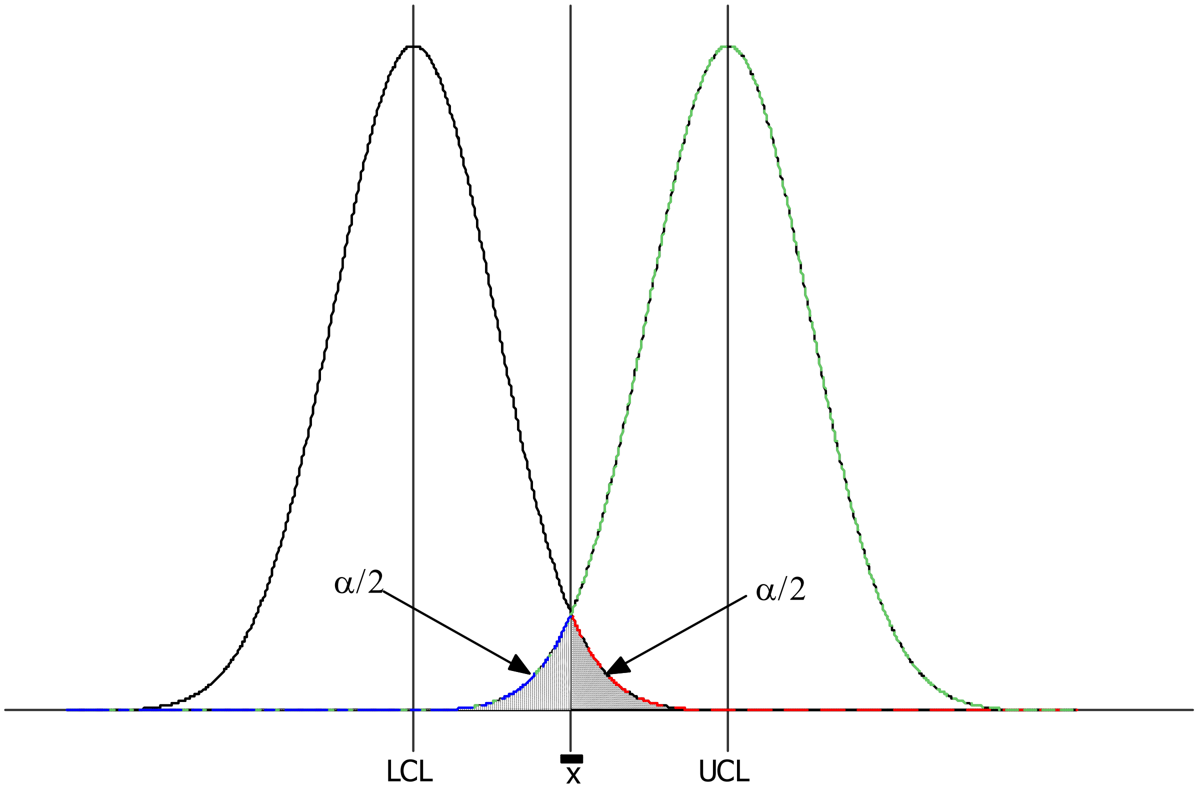
$$\delta = z_{\alpha/2}\sigma_{\bar{x}}$$



Confidence Interval for the Mean

If we solve for μ_x we obtain the confidence interval for the unknown population mean:

$$\Phi(\bar{x} - \delta < \mu_x < \bar{x} + \delta) = 1 - \alpha.$$



Confidence Interval for the Mean

- For a specified value of the confidence interval half-width δ (the *precision* of the estimate) and given (known) σ_x and α related by:

$$\begin{aligned}\delta &= z_{\alpha/2}\sigma_{\bar{x}} \\ &= z_{\alpha/2}\sigma_x/\sqrt{n}\end{aligned}$$

  we can solve for the sample size:

$$n = \left(\frac{z_{\alpha/2}\sigma_x}{\delta}\right)^2.$$

- When the population standard deviation is unknown and must be estimated from the sample data the normal (z) distribution must be replaced with Student's t distribution:

$$n = \left(\frac{t_{\alpha/2}\widehat{\sigma}_x}{\delta}\right)^2.$$

  Because $t_{\alpha/2}$ depends on n through its degrees of freedom value this equation must be solved by iteration.

Confidence Interval for the Mean

Example: Determine the sample size necesary to estimate, with 95% confidence, the mean of a population with precision $\delta = 10$ when $\widehat{\sigma}_x = 20$.

Solution: If we knew σ_x then:

$$n = \left(\frac{z_{0.025}\sigma_x}{\delta}\right)^2 = \left(\frac{1.96 \times 20}{10}\right)^2 = 16.$$

With $n = 16$, $v = 15$, and $t_{0.025} = 2.13$ so

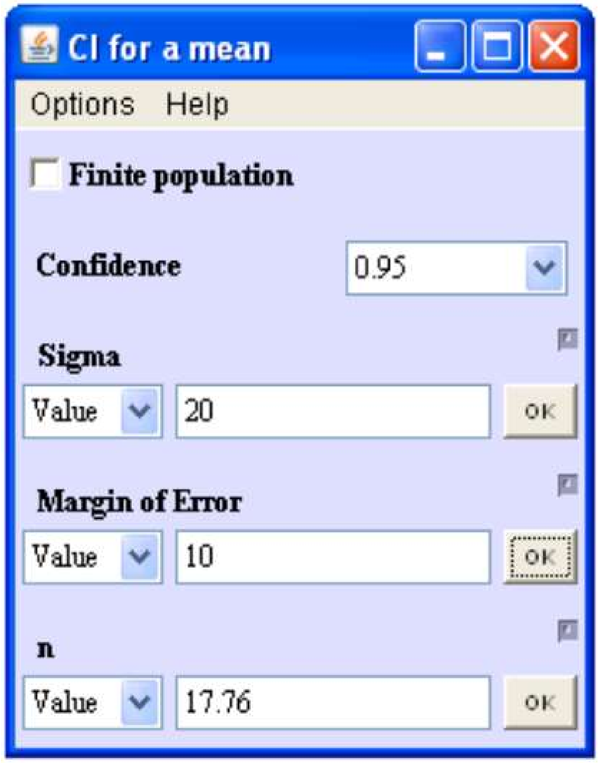
$$n = \left(\frac{t_{0.025}\widehat{\sigma}_x}{\delta}\right)^2 = \left(\frac{2.13 \times 20}{10}\right)^2 = 19.$$

Eventually, with $n = 18$, $v = 17$, and $t_{0.025} = 2.11$:

$$n = \left(\frac{t_{0.025}\widehat{\sigma}_x}{\delta}\right)^2 = \left(\frac{2.11 \times 20}{10}\right)^2 = 18.$$

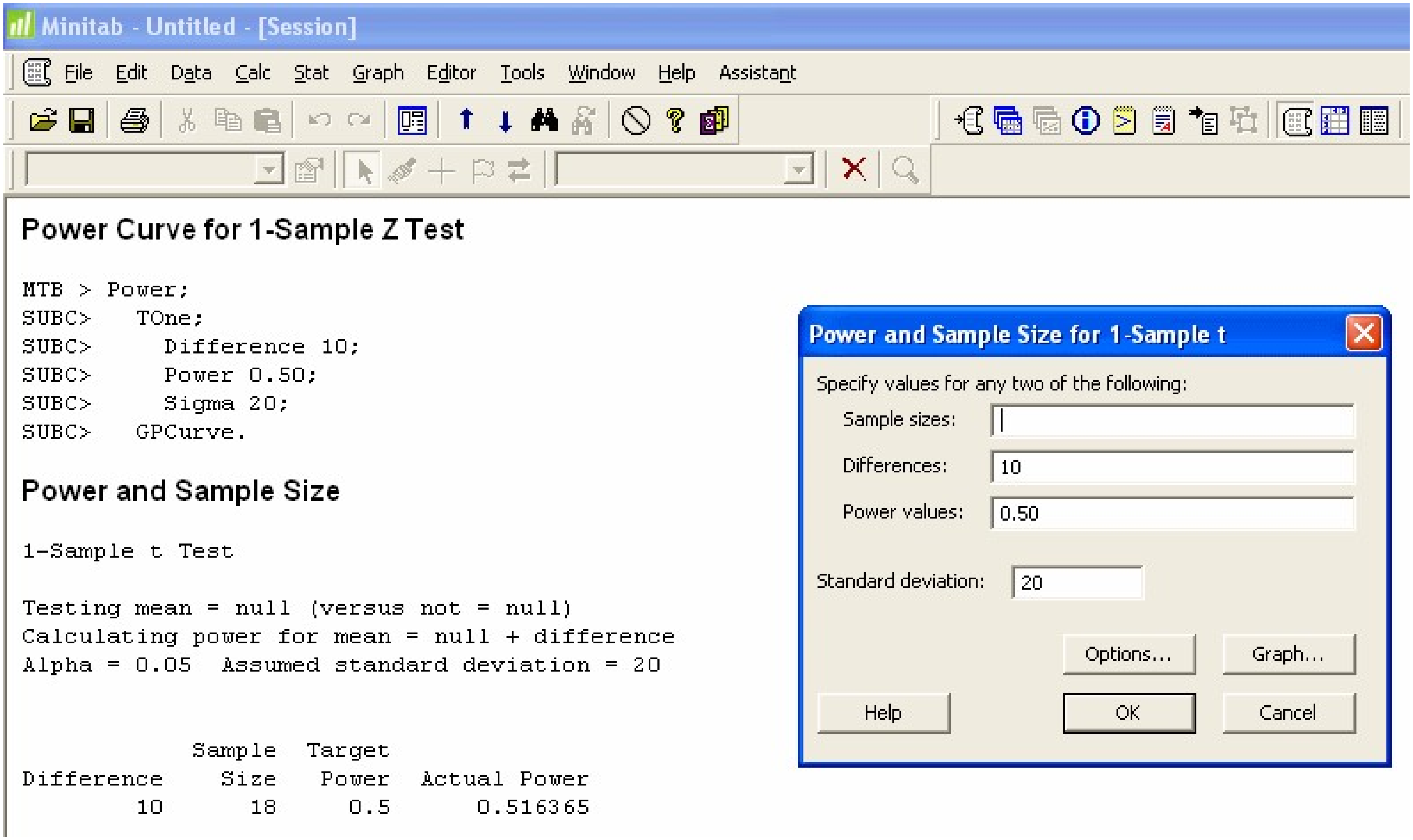
Confidence Interval for the Mean

Solution by Piface with **CI for one mean**:



Confidence Interval for the Mean

Solution by MINITAB with **Stat**> **Power and Sample Size**> **1-Sample t**:



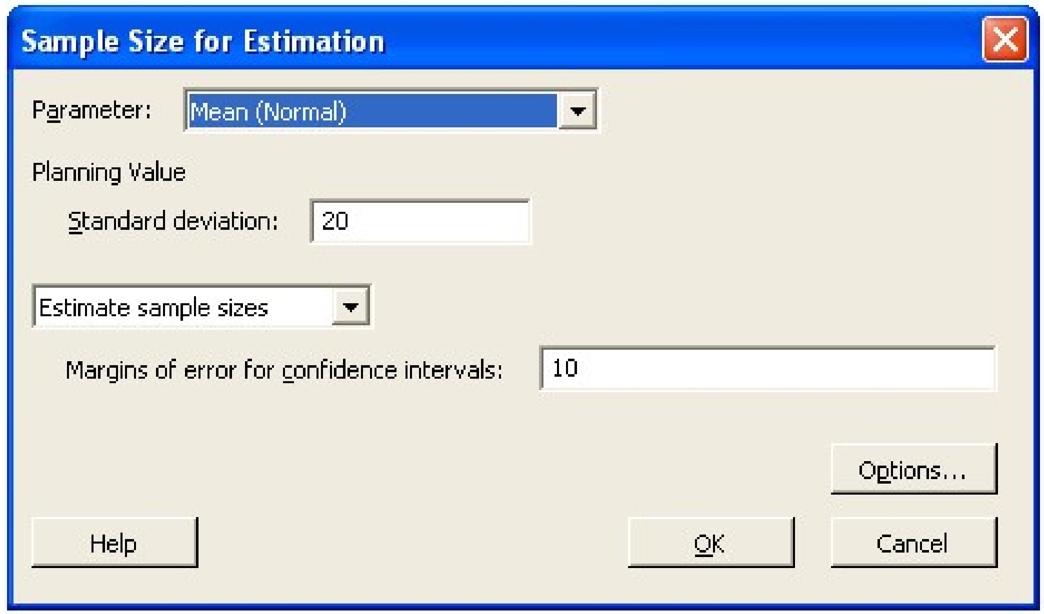
Confidence Interval for the Mean

Solution by MINITAB with **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Mean** (**Normal**):

```
MTB > SSCI;
SUBC>   NMean;
SUBC>     Sigma 20;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 10.
```

Sample Size for Estimation

```
Method

Parameter                Mean
Distribution             Normal
Standard deviation       20 (estimate)
Confidence level         95%
Confidence interval      Two-sided


Results

  Margin   Sample
of Error     Size
      10       18
```



Confidence Interval for the Mean

- To calculate the sample size we need $\alpha, \widehat{\sigma}_x$, and δ.
- Use $\alpha = 0.05$ or whatever value is appropriate.
- Sources for the σ_x estimate:
  - Historical data
  - Preliminary study
  - Data from a similar process
  - Expert opinion
  - Published results (beware of publication bias)
  - Guess
- Confidence interval half-width (δ):
  - Must be chosen by the researcher
  - Must be sufficiently narrow to indicate a unique management action
  - Start from outrageous high and low values, work to the middle
  - Be careful of relative confidence interval half-width

Specifying the Confidence Interval Half-width

- In measurement units:

$$\Phi(\bar{x} - \delta < \mu_x < \bar{x} + \delta) = 1 - \alpha$$

  (Note: This is the only method supported in most sample size calculation software. The other methods express δ in relative terms and are not supported in software.)

- Relative to the mean:

$$\Phi(\bar{x}(1 - \delta) < \mu_x < \bar{x}(1 + \delta)) = 1 - \alpha$$

- Relative to the standard deviation:

$$\Phi(\bar{x} - \delta s < \mu_x < \bar{x} + \delta s) = 1 - \alpha$$

  - Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences.*
  - This method is bad practice! See Russ Lenth’s discussion.

Sensitivity of the Confidence Interval

If the standard deviation is unknown the sample size is

$$n = \left(\frac{t_{\alpha/2}\widehat{\sigma}_x}{\delta} \right)^2$$

- Student's t distribution approaches the normal (z) distribution very quickly so the approximation of $t_{\alpha/2}$ with $z_{\alpha/2}$ has little effect on the sample size unless the sample size is very small.
- Compared to other factors, the magnitude of $t_{\alpha/2}$ or $z_{\alpha/2}$ changes slowly with α so the value of α has little effect on the sample size.
- Sample size is proportional to the square of the standard deviation, i.e. $n \propto \widehat{\sigma}_x^2$, so changes to the estimated value of $\widehat{\sigma}_x$ will have a big effect on sample size. For example, doubling the value of the standard deviation estimate will quadruple the sample size.
- Sample size is inversely proportional to the square of the confidence interval half-width, i.e. $n \propto \frac{1}{\delta^2}$, so changes to the estimated value of δ will have a big effect on sample size.

Sensitivity of the Confidence Interval

- Recommendations:
  - Don't worry too much about the value of α (just use $\alpha = 0.05$).
  - Don't worry too much about the approximation $t_{\alpha/2} \simeq z_{\alpha/2}$.
  - Be very careful determining the standard deviation.
  - Be very careful choosing a value for the confidence interval half-width.

Confidence Interval for the Mean

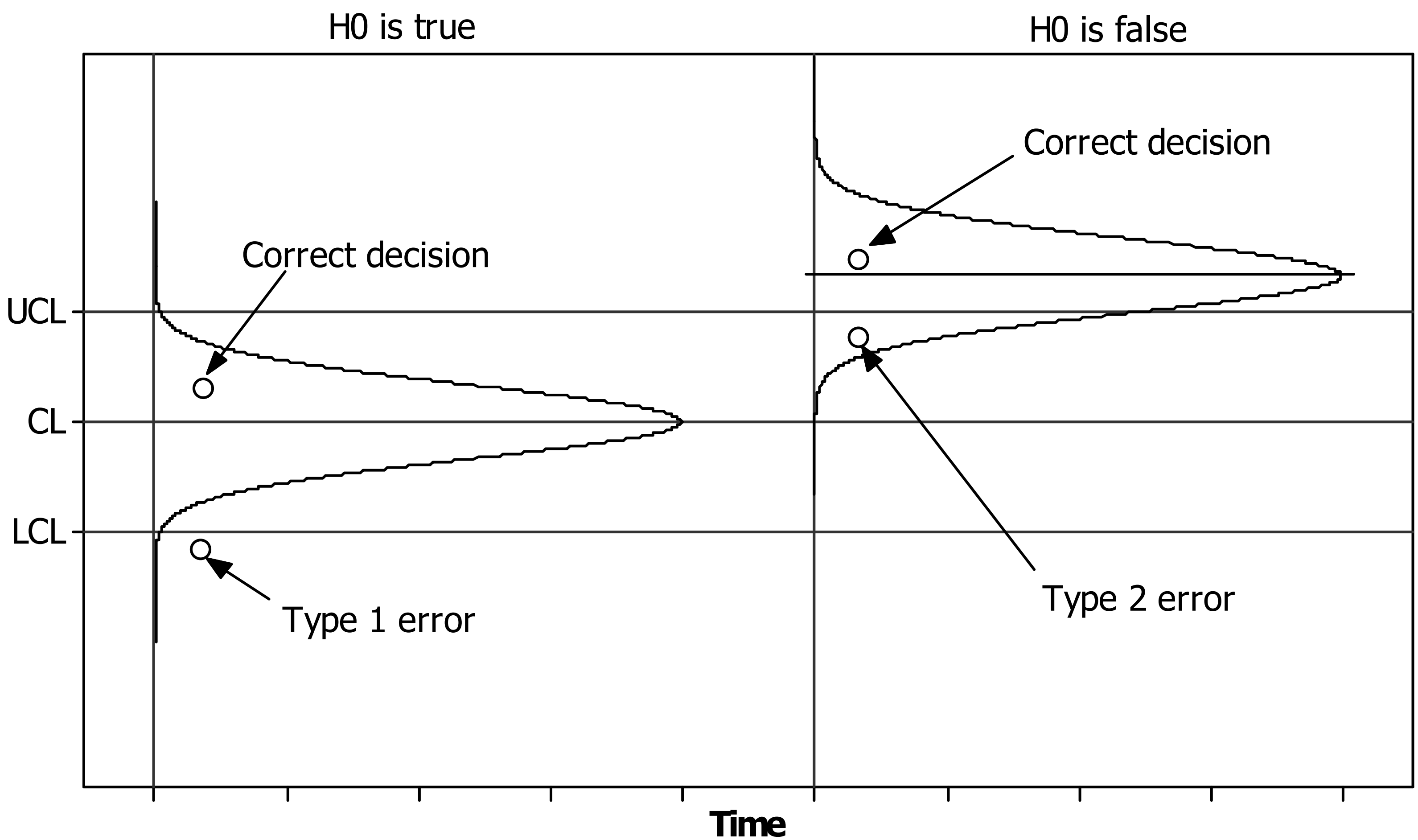
Exercise: How many observations are required to determine the number of flexing cycles required to break paper clips?

- What's easy to determine?
- What's hard to determine?
- What are the consequences of variation in those choices?

Hypothesis Tests

- Hypothesis tests are used to test and compare population parameters and distribution shapes of one, two, and many populations.
- Hypothesis tests involve two complementary hypotheses: the null hypothesis (H_0) and the alternate hypothesis (H_A).
- State what's to be demonstrated in H_A and its complement is H_0.
- *"The extraordinary claim requires extraordinary evidence."* - Carl Sagan. The extraordinary claim is H_A and its complement, the status quo, is H_0.
- Reject H_0 in favor of H_A when the sample data are statistically unlikely to occur under H_0.
- There is no opportunity to accept H_0.

Errors in Hypothesis Testing



Test for the Mean (σ Known)

- The hypotheses to be tested are:

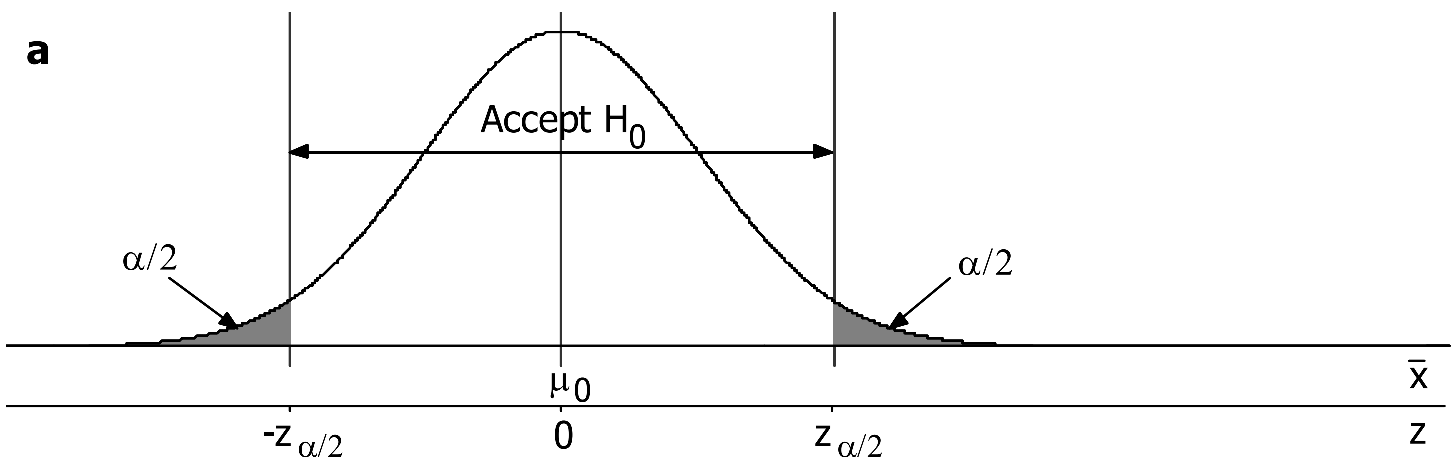
$$H_0 : \mu = \mu_0 \text{ versus } H_A : \mu \neq \mu_0.$$

- The test statistic is:

$$z = \frac{\bar{x} - \mu_0}{\sigma_x / \sqrt{n}}.$$

- The acceptance interval for H_0 is:

$$\Phi(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha.$$



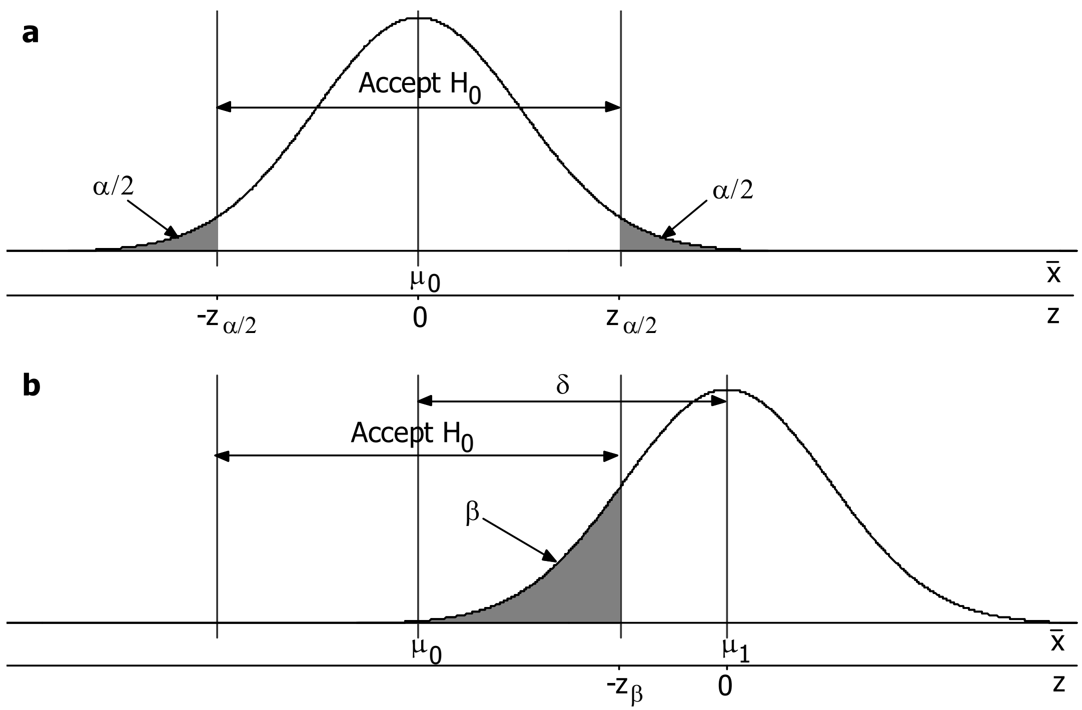
Test for the Mean

- The effect size δ is:

$$\delta = \mu_1 - \mu_0 = (z_{\alpha/2} + z_\beta)\frac{\sigma_x}{\sqrt{n}}$$

- The power of the test is:

$$\pi = \Phi(-z_\beta < z < \infty) \text{ where } z_\beta = \frac{\delta}{\sigma_x/\sqrt{n}} - z_{\alpha/2}$$



Test for the Mean

- The quantity:

$$\phi = z_{\alpha/2} + z_\beta = \frac{\delta}{\sigma_x/\sqrt{n}}$$

  is called the *noncentrality parameter*.
- The sample size is:

$$n = \left((z_{\alpha/2} + z_\beta) \frac{\sigma_x}{\delta} \right)^2$$

- The values of α and β have little effect on the sample size so, as with confidence intervals, focus your attention on determining appropriate values for σ_x and δ.

Test for the Mean

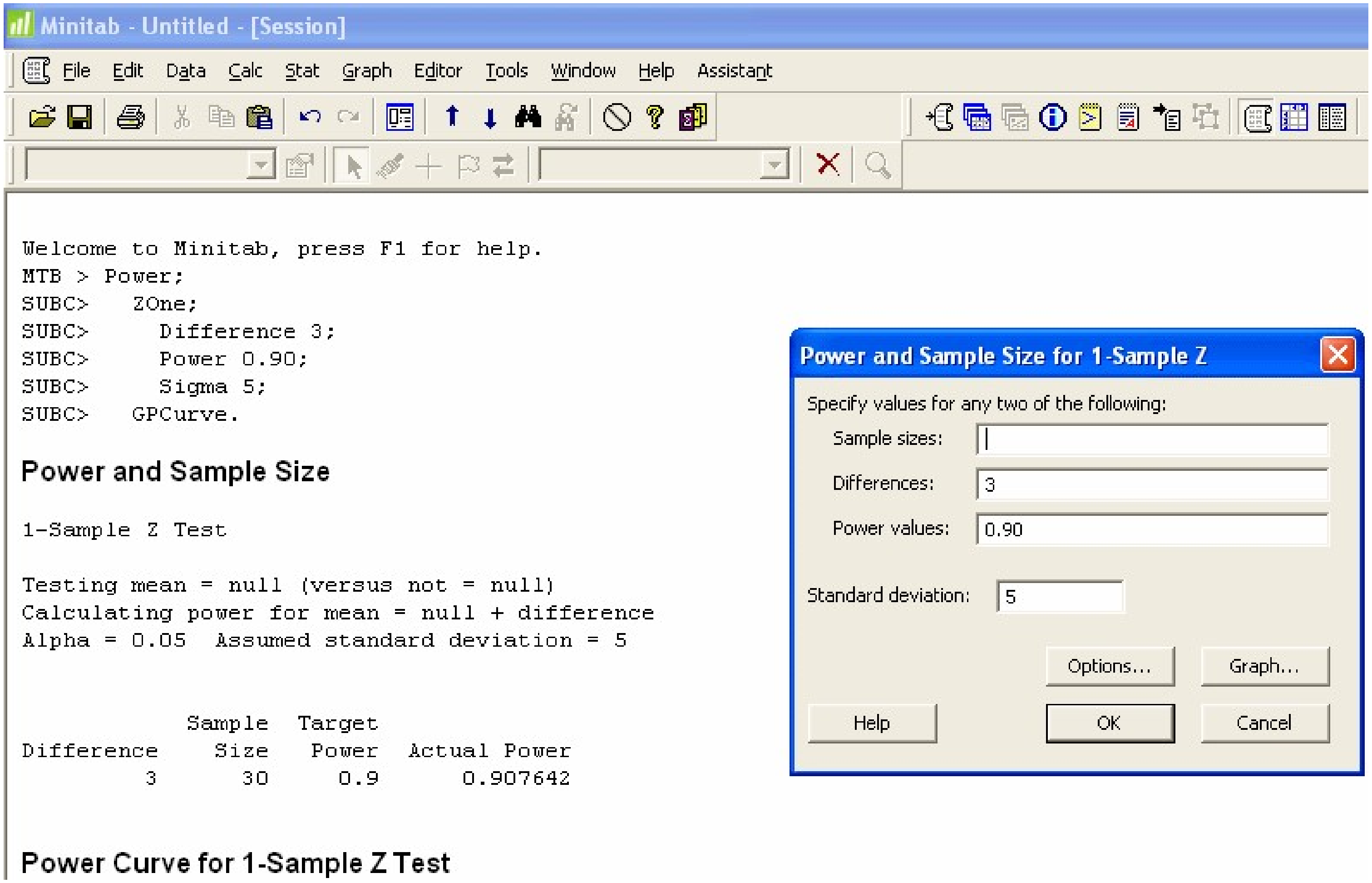
Example: An experiment will be performed to test $H_0 : \mu = 60$ versus $H_A : \mu \neq 60$. The standard deviation is known to be $\sigma_x = 5$. What sample size is required to reject H_0 with $\pi = 0.90$ when $\mu = 63$? Assume that the distribution of x is normal and use $\alpha = 0.05$.

Solution: The effect size is $\delta = 63 - 60 = 3$ with associated power $\pi = 0.90$ or type 2 error rate $\beta = 0.10$. Then the sample size must be:

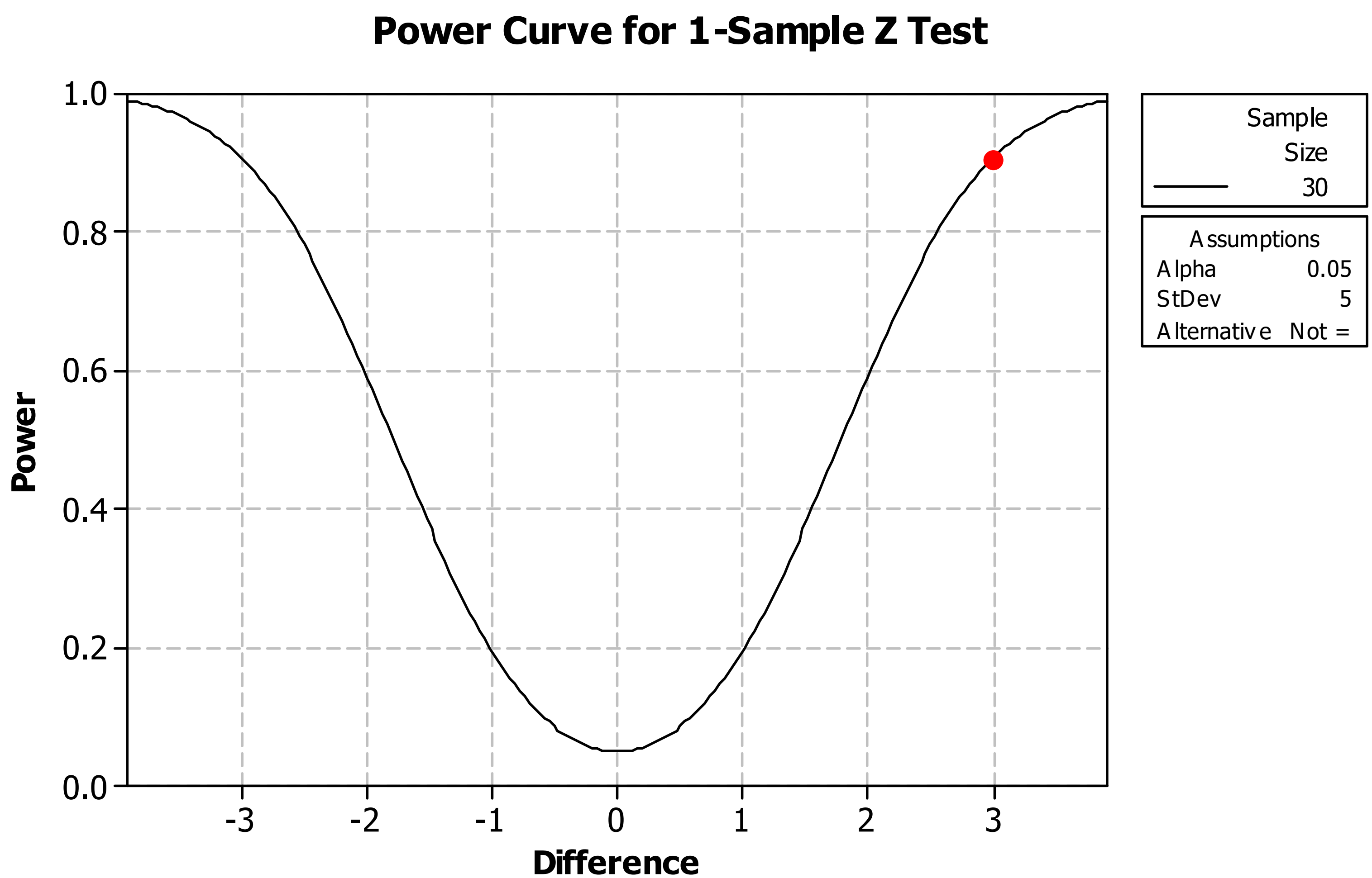
$$
\begin{aligned}
n &= \left(\frac{(z_{0.025} + z_{0.10})\sigma_x}{\delta} \right)^2 \\
&= \left(\frac{(1.96 + 1.282)5}{3} \right)^2 \\
&= 30
\end{aligned}
$$

Test for the Mean

Solution by MINITAB with **Stat**> **Power and Sample Size**> **1-Sample Z**:



Test for the Mean



Test for the Mean (σ_x Unknown)

- The hypotheses to be tested are:

$$H_0 : \mu = \mu_0 \text{ versus } H_A : \mu \neq \mu_0.$$

- If σ_x is unknown, the test statistic is:

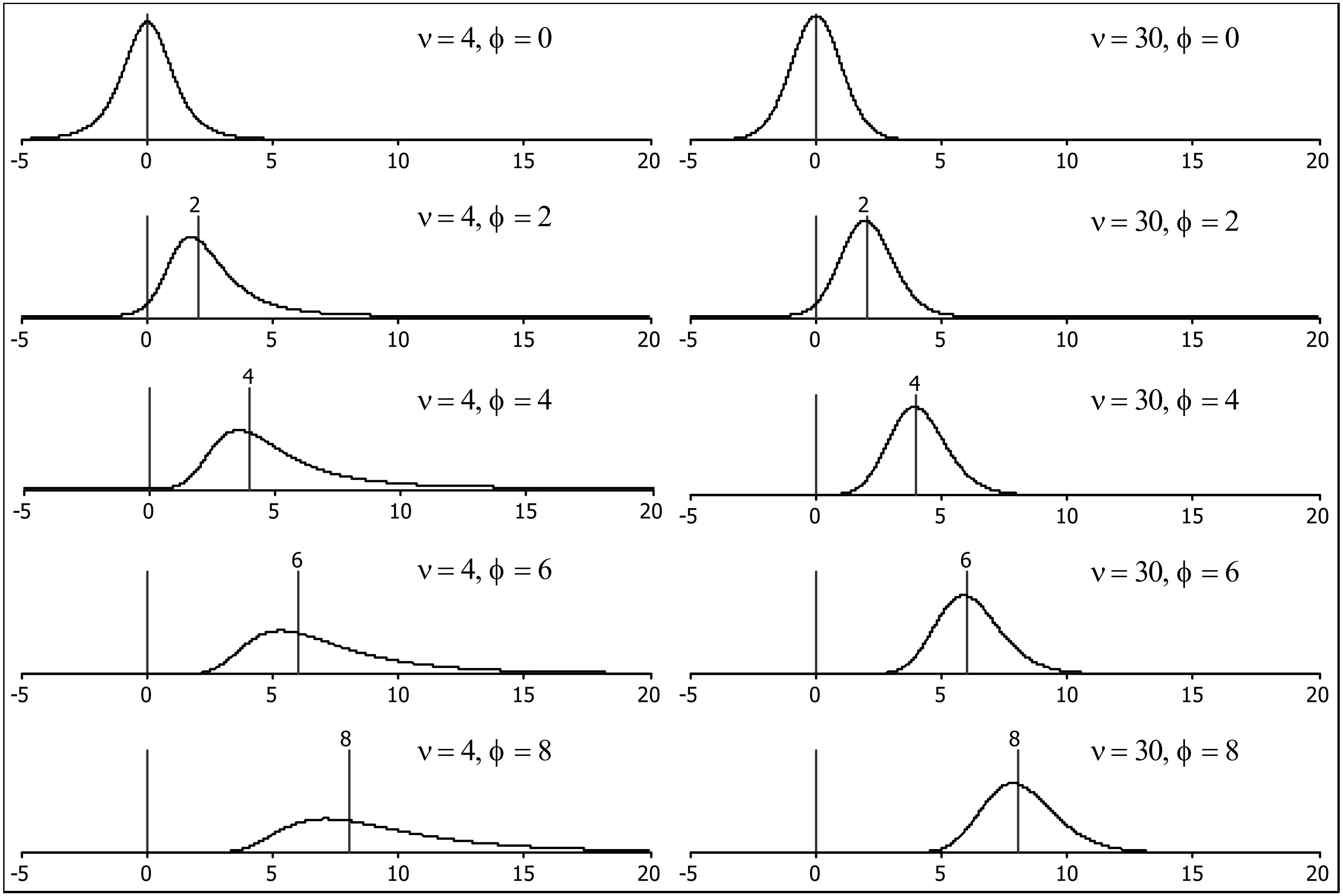
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

- The acceptance interval for H_0 is:

$$\Phi(-t_{\alpha/2,\nu} < t < t_{\alpha/2,\nu}) = 1 - \alpha \text{ where } \nu = n - 1.$$

- The distribution of t under H_0 is the well known *central* Student's t distribution but the distribution of t under H_A is the less well known *noncentral* Student's t distribution.

Test for the Mean (σ_x Unknown)



Test for the Mean (σ_x Unknown)

- The exact power is given by:

$$t_{\alpha/2} = t_{\beta,\phi} \text{ where } \phi = \frac{\delta}{\hat{\sigma}_x/\sqrt{n}}.$$

- Find the sample size for a target power value by iterating over the sample size until the power requirement is satisfied.
- When the sample size is large, $t \simeq z$ and the normal distribution power and sample size methods are good approximations.
- Use $t \simeq z$ as a starting point for manual iterations to find a sample size.

Test for the Mean (σ_x Unknown)

Example: For the one-sample test of $H_0 : \mu = 30$ versus $H_A : \mu \neq 30$, what sample size is required to detect a shift to $\mu = 32$ with 90% power? The population standard deviation is unknown but expected to be $\sigma \simeq 1.5$.

Test for the Mean (σ_x Unknown)

Solution: Assuming that $t \simeq z$:

$$n = \left(\frac{(z_{0.025} + z_{0.10})\sigma_x}{\delta} \right)^2$$

$$= \left(\frac{(1.96 + 1.282)1.5}{2} \right)^2$$

$$= 6$$

Try again:

$$n = \left(\frac{(t_{0.025,5} + t_{0.10,5})\widehat{\sigma}_x}{\delta} \right)^2$$

$$= \left(\frac{(2.57 + 1.48)1.5}{2} \right)^2$$

$$= 9.2$$

Further iterations indicate that $n = 9$.

Test for the Mean (σ_x Unknown)

Solution (**continued**): The *exact* power with $n = 9$ is given by the solution to:

$$t_{\alpha/2} = t_{\beta,\phi} \text{ where } \phi = \frac{\delta}{\widehat{\sigma}_x/\sqrt{n}}.$$

So

$$\phi = \frac{2}{1.5/\sqrt{9}} = 4.0$$

and

$$t_{0.025} = 2.306 = t_{\beta,4.0}$$

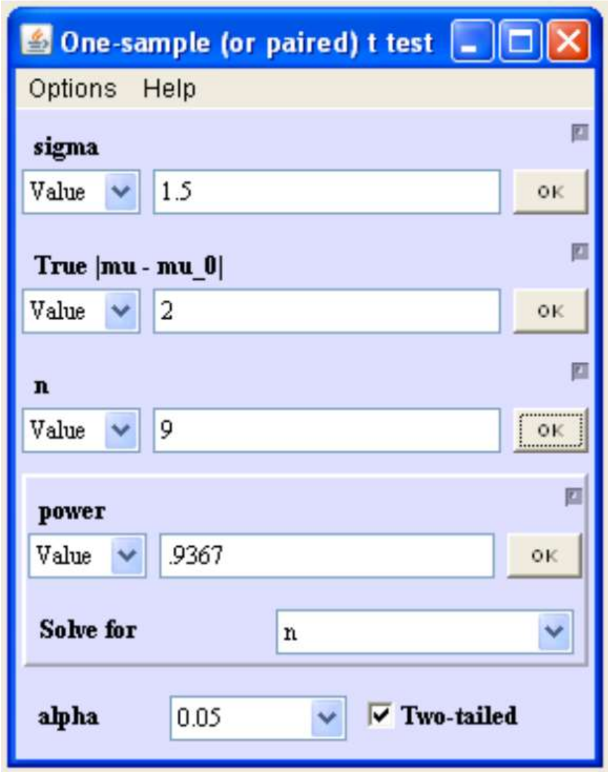
which gives

$$\pi = 1 - \beta = 0.9366$$

(Hint: In MINITAB use the **Calc**> **Probability Distributions**> **t** menu to perform the calculations.)

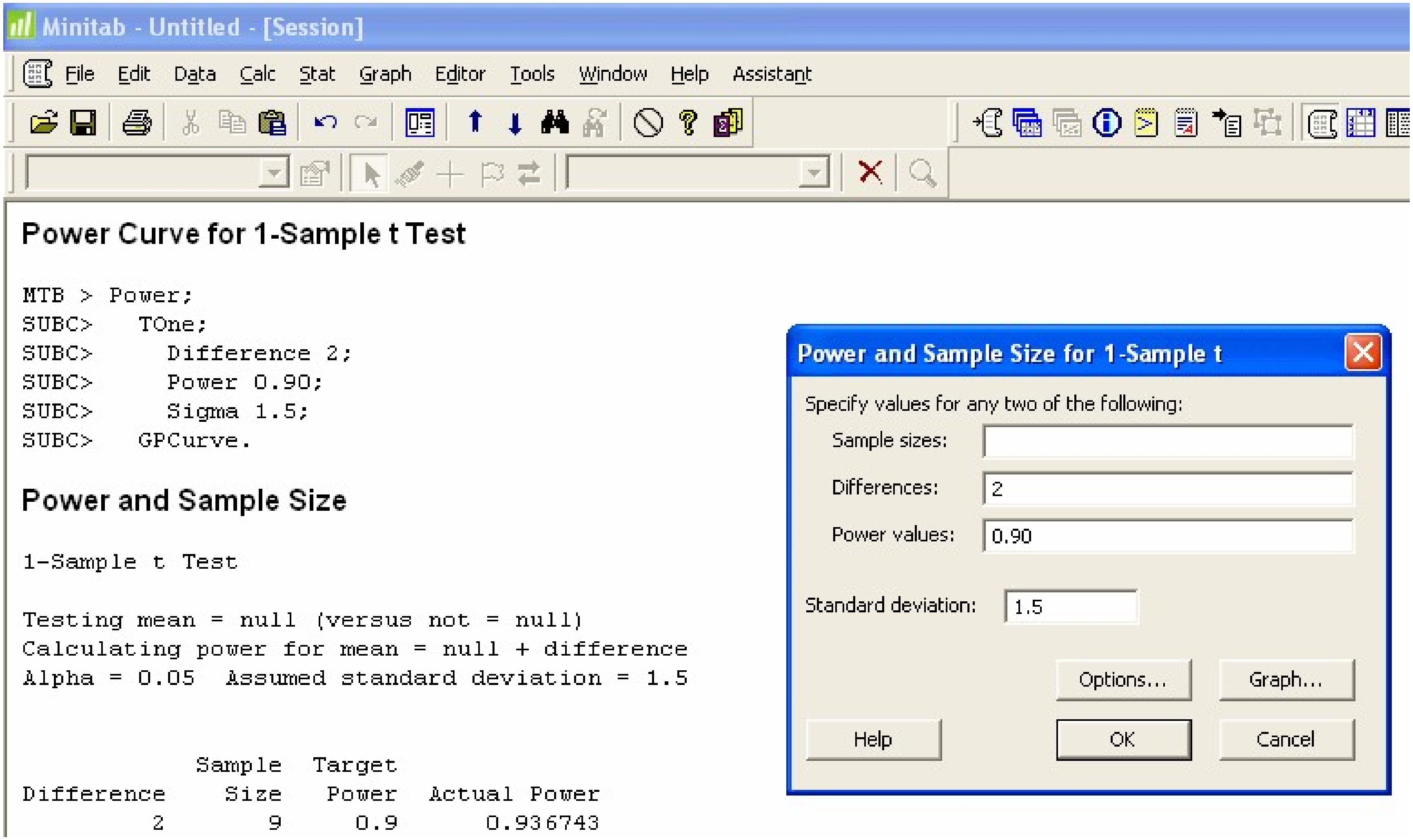
Test for the Mean (σ_x Unknown)

Solution by Piface with **One-sample t test** (**or paired t**):



Test for the Mean (σ_x Unknown)

Solution by MINITAB with **Stat**> **Power and Sample Size**>
1-Sample t:



Test for Two Means

There are too many variations on tests for two means to discuss all of them here. The conditions that affect which test to use are:

- Known or unknown standard deviations
- Equal or unequal standard deviations
- Equal or unequal sample sizes
- One sample size fixed
- Significance test or equivalence test

Test for Two Means

In the test of $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$ when the standard deviations are known and equal (i.e., $\sigma_1 = \sigma_2 = \sigma_\epsilon$), the test statistic is:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_\epsilon \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

When $n_1 = n_2 = n$, the power is given by

$$\pi = P(-\infty < z < z_\beta)$$

where

$$z_\beta = \sqrt{\frac{n}{2}} \frac{\Delta\mu}{\widehat{\sigma}_\epsilon} - z_{\alpha/2}.$$

The approximate sample size to obtain a specified power value is

$$n_1 = n_2 = 2(z_{\alpha/2} + z_\beta)^2 \left(\frac{\widehat{\sigma}_\epsilon}{\Delta\mu} \right)^2.$$

Test for Two Means

In the test of $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$ when the treatments are homoscedastic (i.e., $\sigma_1 = \sigma_2 = \sigma_\epsilon$) but the standard deviations are unknown the test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_\epsilon \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where s_ϵ is the standard error. When $n_1 = n_2 = n$, the power is given by

$$\pi = P(-\infty < t < t_\beta)$$

where

$$t_\beta = \sqrt{\frac{n}{2}} \frac{\Delta\mu}{\widehat{\sigma}_\epsilon} - t_{\alpha/2}.$$

The approximate sample size to obtain a specified power value is

$$n_1 = n_2 = 2(t_{\alpha/2} + t_\beta)^2 \left(\frac{\widehat{\sigma}_\epsilon}{\Delta\mu} \right)^2.$$

Test for Two Means

Example: An experiment will be performed to test $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$. The standard deviations are homoscedastic and known to be $\sigma_x = 50$. What sample size is required to reject H_0 with $\pi = 0.90$ when the difference between the means is $\Delta\mu = 80$? Assume that the distributions are normal and use $\alpha = 0.05$.

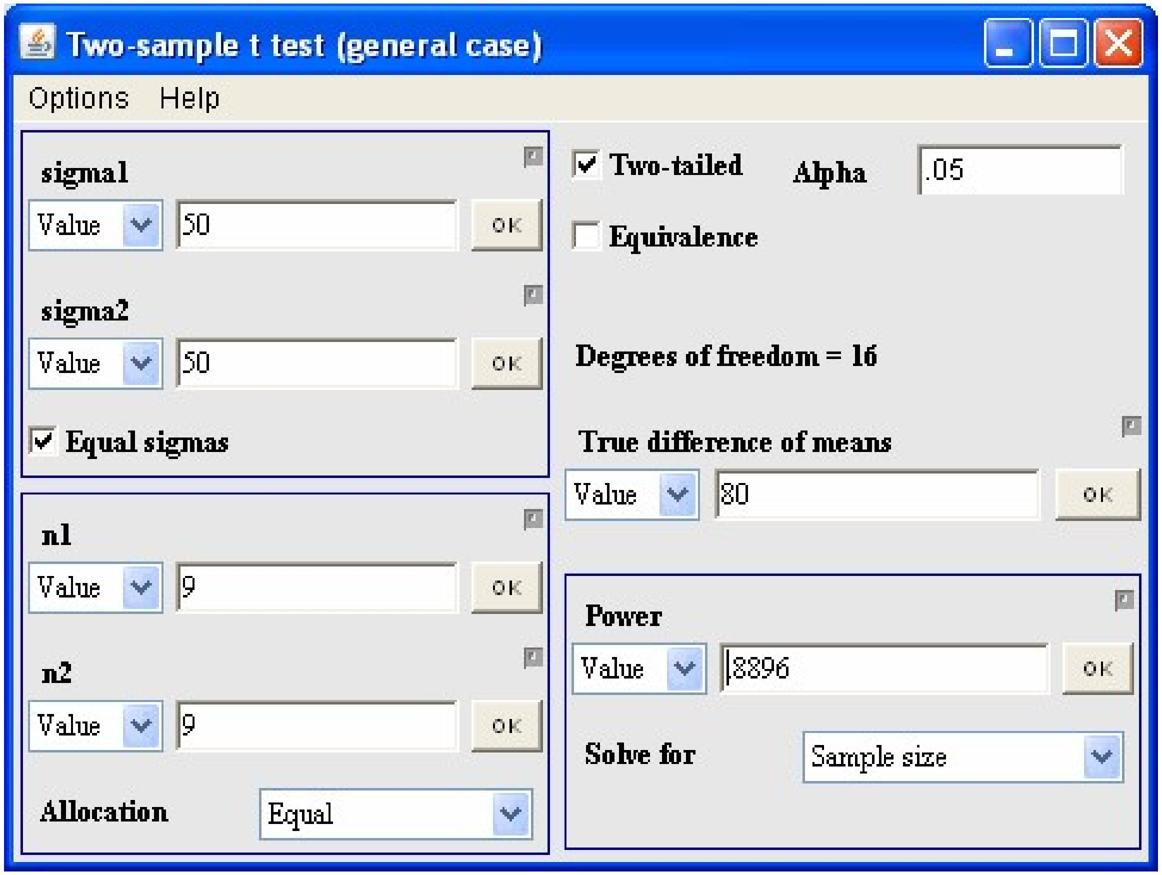
Solution: Starting from $t \simeq z$:

$$\begin{aligned} n &= 2(z_{0.025} + z_{0.10})^2 \left(\frac{\sigma_\epsilon}{\Delta\mu} \right)^2 \\ &= 2(1.96 + 1.282)^2 \left(\frac{50}{80} \right)^2 \\ &= 9 \end{aligned}$$

Another iteration using t gives $n = 10$.

Test for Two Means

Solution by Piface with **Two-sample t test**:



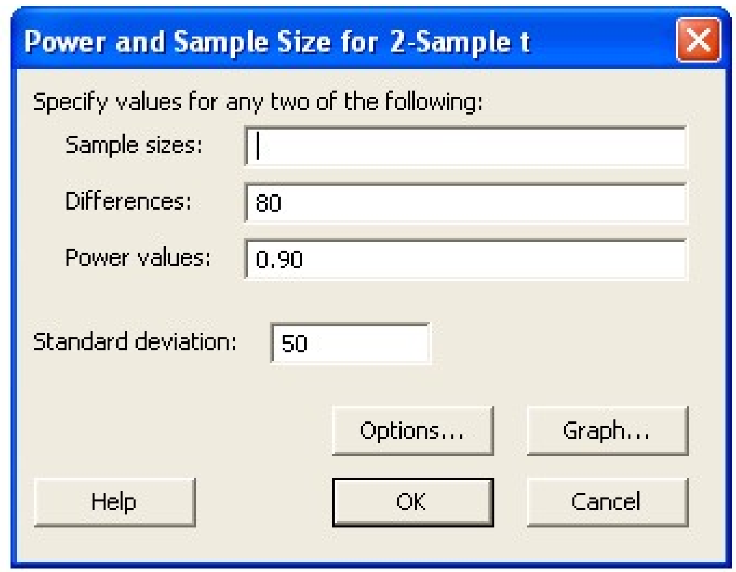
Test for Two Means

Solution by MINITAB **Stat**> **Power and Sample Size**> **2-Sample t**:

```
MTB > Power;
SUBC>   TTwo;
SUBC>     Difference 80;
SUBC>     Power 0.90;
SUBC>     Sigma 50;
SUBC>   GPCurve.
```

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 50


            Sample  Target
Difference    Size   Power  Actual Power
        80      10     0.9      0.922373

The sample size is for each group.
```



Multiple Comparisons Tests

- If simultaneous confidence intervals or tests are required, use a *multiple comparisons* method to control the type I error rate for the *family* of tests. For example:
  - Bonferroni-corrected two-sample *t* tests
  - Tukey's HSD test for all possible comparisons.
  - Dunnett's test for comparisons between treatments and a control.
  - Hsu's test for comparisons to the best (highest or lowest).

Multiple Comparisons Tests

- Apply *Bonferroni's correction* to the type 1 error rate and use the relevant two-sample test to approximate power and sample size for multiple comparisons tests as long as the number of comparisons isn't too large. In general, if there are K simultaneous confidence intervals or tests planned then α for the individual confidence intervals or tests should be:

$$\alpha = \frac{\alpha_{family}}{K}.$$

- Bonferroni's method is conservative,i.e. insensitive to small but possibly important differences between the treatment means, so the sample size calculated using the Bonferroni correction will be slightly larger than the exact sample size calculated for other analysis methods such as Tukey's and Dunnett's. The difference between the approximate and exact sample sizes is usually small compared to the effects on sample size caused by uncertainties in the values of the standard error and the effect size.

Multiple Comparisons Tests

- When the number of multiple comparisons becomes very large Bonferroni's method becomes very conservative. A less conservative method of correcting α for individual tests is Sidak's or Dunn's method:

$$\alpha = 1 - (1 - \alpha_{family})^{\frac{1}{K}}$$

where K is the number of tests required.

Multiple Comparisons Tests

Bonferroni's method applied to all possible pairwise comparisons among k treatments leads to

$$K = \binom{k}{2} = \frac{k(k-1)}{2}$$

and

$$n_i = 2(z_{\alpha/2} + z_\beta)^2 \left(\frac{\sigma_\epsilon}{\Delta\mu} \right)^2$$

where

$$\alpha = \frac{\alpha_{family}}{K}.$$

Note: The sample size given by this method is very close to the exact sample size calculated for ANOVA.

Multiple Comparisons Tests

Example: For five treatment groups, determine the sample size required per treatment to detect a difference $\Delta\mu = 200$ between two treatment means using Bonferroni-corrected two-sample t tests for all possible pairs of treatments with 90% power. Assume that the five populations are normal and homoscedastic with $\widehat{\sigma}_\epsilon = 100$.

Multiple Comparisons Tests

Solution: With $k = 5$ treatments there will be $K = \binom{5}{2} = 10$ two-sample t tests to perform. To restrict the family error rate to $\alpha_{family} = 0.05$, the Bonferroni-corrected error rate for individual tests is

$$\alpha = \frac{0.05}{10} = 0.005.$$

With $t \simeq z$, the sample size per treatment group is

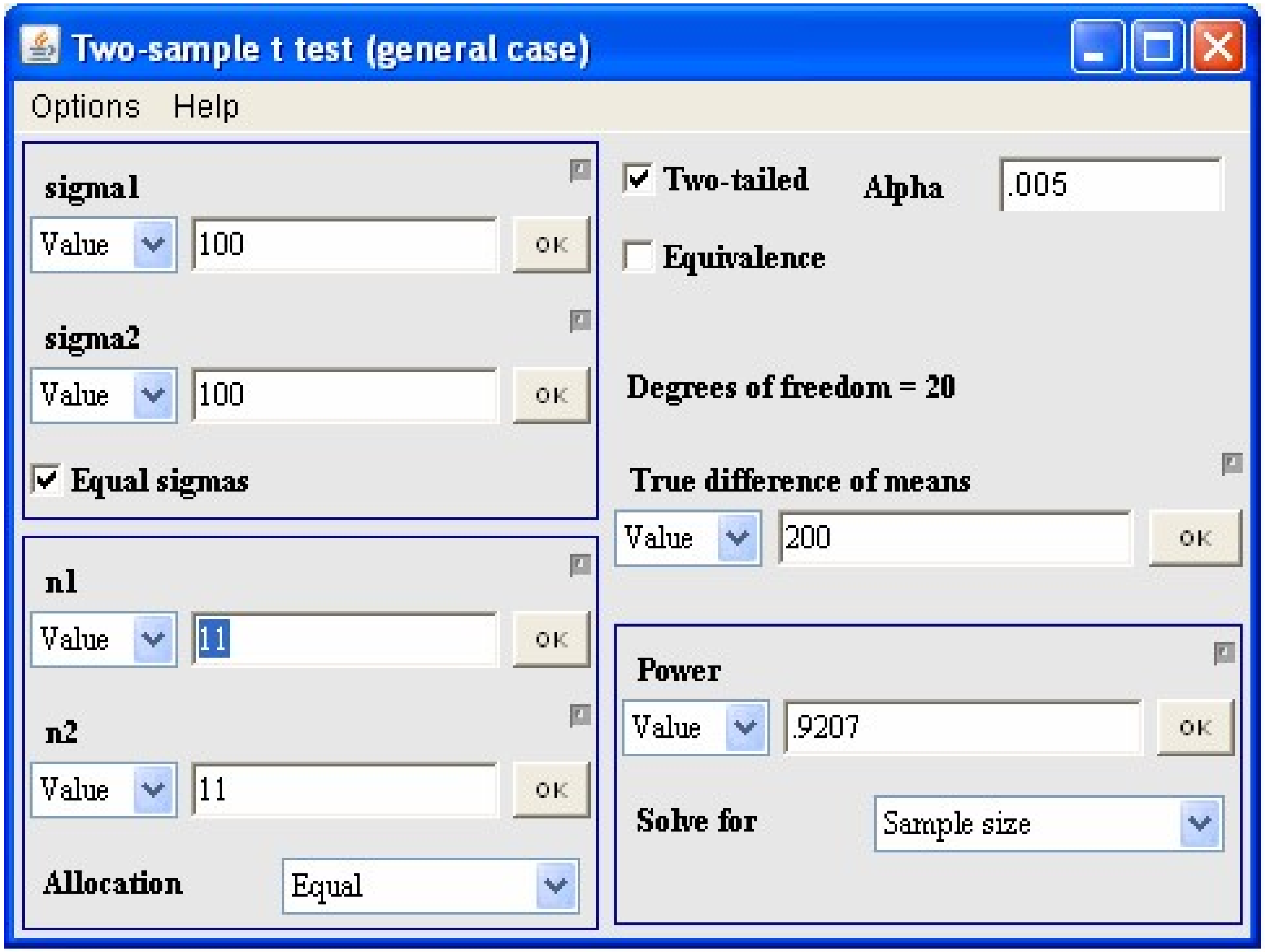
$$n_i = 2\left(\frac{(z_{0.0025} + z_{0.10})\widehat{\sigma}_{\epsilon}}{\Delta\mu}\right)^2$$

$$= 2\left(\frac{(2.81 + 1.282)100}{200}\right)^2 = 9.$$

Further iterations using t instead of z converge to $n = 12$.

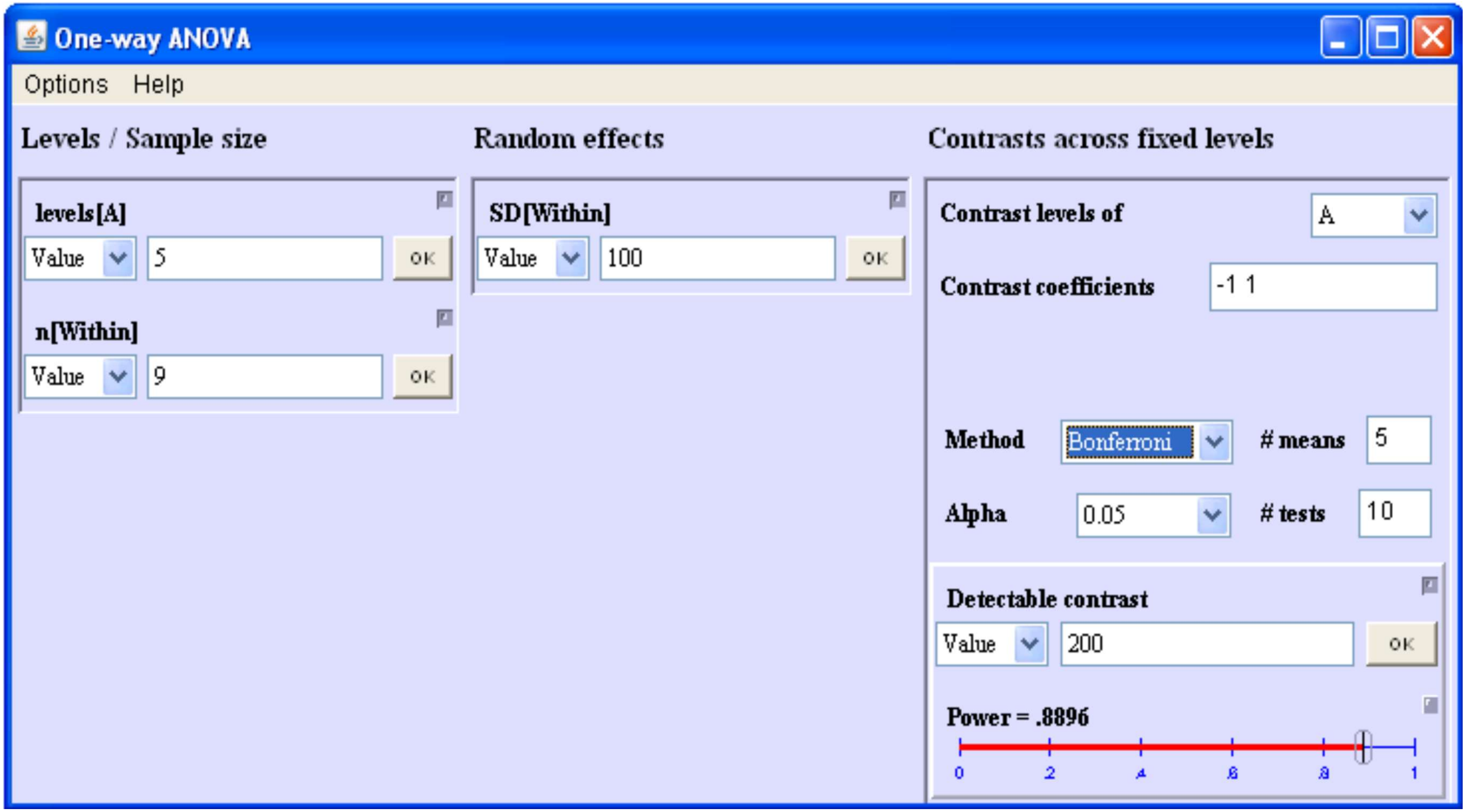
Multiple Comparisons Tests

Solution by Piface:



Multiple Comparisons Tests

Solution by Piface:



Multiple Comparisons Tests

In an experiment to compare K treatment groups to a single control group (that is, comparisons between treatments are not of interest), the control group's importance in the analysis suggests that it deserves to have a larger sample size than the treatment groups. The near-optimal allocation of observations to treatment and control groups is

$$n_0 = n_i \sqrt{K}$$

where n_0 is the number of observations in the control group and n_i is the number observations in each of the treatment groups. The sample size for the treatment groups is given by:

$$n_i = \left(1 + \frac{1}{\sqrt{K}}\right)(z_{\alpha/2} + z_\beta)^2 \left(\frac{\sigma_\epsilon}{\Delta\mu}\right)^2$$

where the Bonferroni-corrected type 1 error rate is

$$\alpha = \frac{\alpha_{family}}{K}.$$

Multiple Comparisons Tests

Example: Determine the sample size required to test five treatment groups against a control group if the tests must detect a difference $\Delta\mu = 200$ between a treatment mean and the control group mean. Assume near-optimal allocation of samples to the treatments and the control and 90% power. Assume that all of the populations are normal and homoscedastic with $\widehat{\sigma}_{\epsilon} = 100$.

Multiple Comparisons Tests

Solution: With five treatment groups and one control group there will be $K = 5$ two-sample t tests. To restrict the family error rate to $\alpha_{family} = 0.05$ the Bonferroni-corrected error rate for individual tests is

$$\alpha = \frac{0.05}{5} = 0.01.$$

With $t \simeq z$, the sample size for the treatment groups will be

$$n_i = \left(1 + \frac{1}{\sqrt{K}}\right)\left(\frac{(z_{0.005} + z_{0.10})\widehat{\sigma}_\epsilon}{\Delta\mu}\right)^2$$

$$= \left(1 + \frac{1}{\sqrt{5}}\right)\left(\frac{(2.575 + 1.282)100}{200}\right)^2 = 6$$

and the sample size for the control group will be

$$n_0 = n_i\sqrt{K} = 6\sqrt{5} = 14.$$

The number of error degrees of freedom will be $df_\epsilon = 5(6-1) + (14-1) = 38$ so the assumption that $t \simeq z$ is justified.

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. **Standard Deviations**
    a. **One Standard Deviation**
    b. **Pilot Studies**
    c. **Two Standard Deviations**
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Standard Deviations

- Sample size and power calculations for standard deviations are based on the chi-square (χ^2) distribution.
- The accuracy of the χ^2 distribution is VERY sensitive to deviations from normality so be very careful to check the normality assumption.
- The χ^2 distribution can be difficult to work with but it can be approximated by the normal distribution when the sample size is sufficiently large.
- Follow up an approximate sample size calculation with an exact calculation if the approximate method delivers a small sample size.

Confidence Interval for σ

- If the population being sampled is normal, the distribution of $(n-1)s^2/\sigma^2$ is χ^2 with $\nu = n-1$ degrees of freedom, so:

$$P\left(\chi^2_{\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right) = 1-\alpha.$$

- The exact confidence interval for σ is:

$$P\left(s\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}} < \sigma < s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}}\right) = 1-\alpha.$$

Confidence Interval for σ

- When the sample size is very large, the χ^2 distribution is approximately normal with $\mu_{\chi^2} = v$ and $\sigma_{\chi^2} = \sqrt{2v}$.
- Then an approximate large sample confidence interval is:

$$P\left(v - z_{\alpha/2}\sqrt{2v} < \chi^2 < v + z_{\alpha/2}\sqrt{2v}\right) = 1 - \alpha$$

$$\cdots = 1 - \alpha$$

$$P(s(1-\delta) < \sigma < s(1+\delta)) = 1 - \alpha$$

where

$$\delta = \frac{z_{\alpha/2}}{\sqrt{2n}}.$$

- The sample size required to obtain a specified relative confidence interval half-width is

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$

Confidence Interval for σ

Example: What sample size is required to estimate σ with 10% precision and 95% confidence?

Solution: The desired confidence interval has the form

$$P(s(1-0.10) < \sigma < s(1+0.10)) = 0.95.$$

With $\alpha = 0.05$ and $\delta = 0.10$, the sample size required to obtain a confidence interval of the desired half-width is

$$n = \frac{1}{2}\left(\frac{1.96}{0.10}\right)^2 = 193.$$

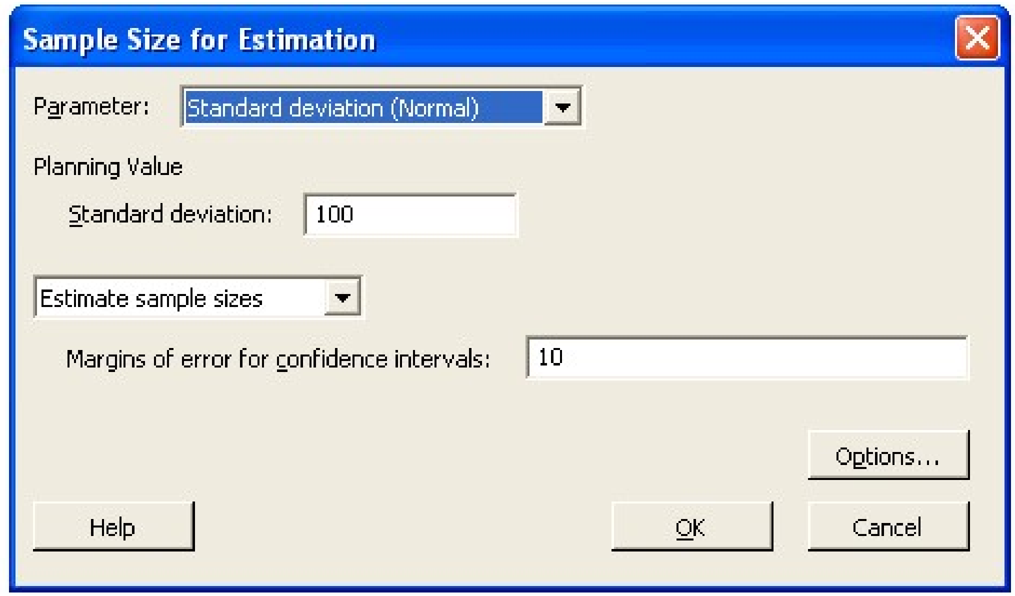
Confidence Interval for σ

Solution by MINITAB using **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Standard Deviation** (**Normal**):

```
MTB > SSCI;
SUBC>   NStDev 100;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 10.
```

Sample Size for Estimation

```
Method

Parameter             Standard deviation
Distribution          Normal
Standard deviation    100
Confidence level      95%
Confidence interval   Two-sided


Results

  Margin  Sample
of Error    Size
      10     234
```



Pilot Studies

- What sample size is required for a pilot study to obtain a sufficiently accurate estimate of the standard deviation to use in the sample size calculation for a primary experiment?
- If δ is the maximum allowable relative error in the sample size of the primary experiment with associated confidence level $1 - \alpha$, that is,

$$P(\hat{n}(1-\delta) < n < \hat{n}(1-\delta)) = 1 - \alpha,$$

then the sample size of the pilot study must be:

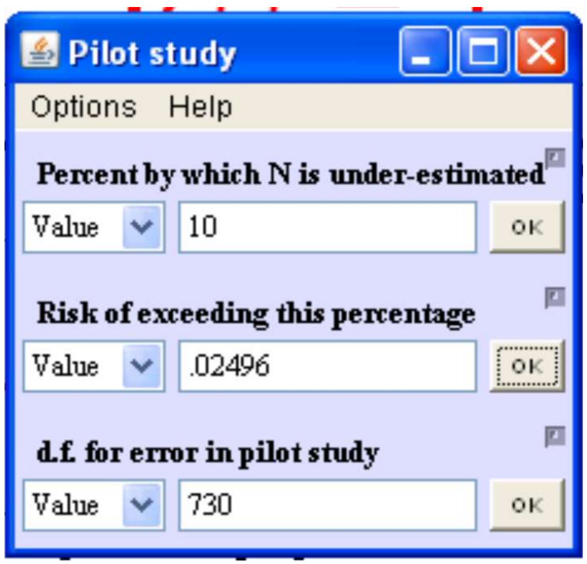
$$n \simeq 2\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$

Pilot Studies

Example: The sample size required in a preliminary experiment to determine σ_x sufficiently well so that the sample size in a primary experiment is within 10% of the correct value with 95% confidence is:

$$n \simeq 2\left(\frac{1.96}{0.1}\right)^2$$
$$\simeq 769$$

Using Piface **Pilot study**:



Test for σ^2

- The hypotheses to be tested are:

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_A : \sigma^2 > \sigma_0^2.$$

- The test statistic is:

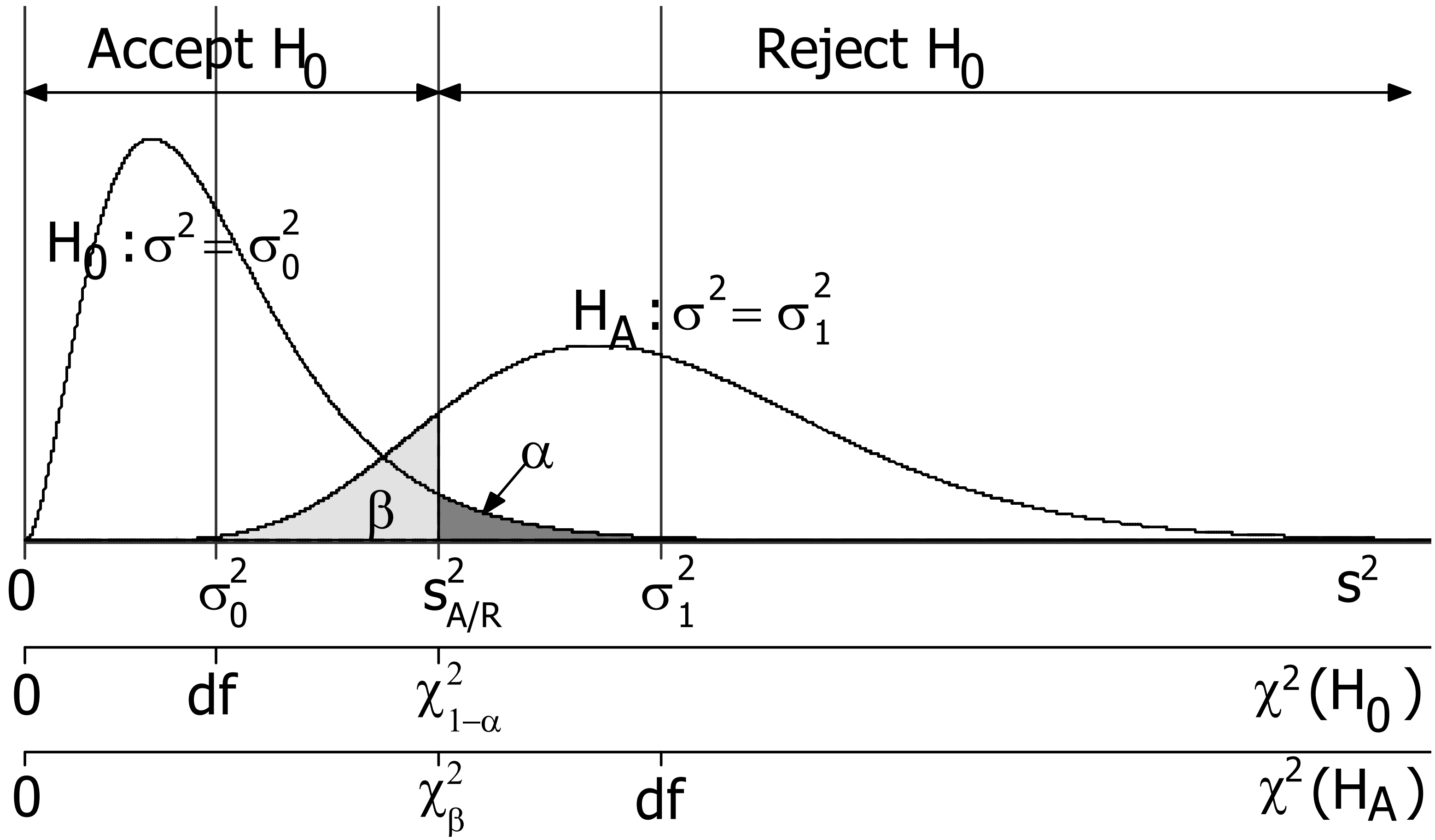
$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

where the χ^2 distribution has $\nu = n - 1$ degrees of freedom.

Test for σ^2

- At the critical value of the sample variance $s^2_{A/R}$:

$$s^2_{A/R} = \frac{\chi^2_{1-\alpha}\sigma^2_0}{n-1} = \frac{\chi^2_\beta \sigma^2_1}{n-1}$$



Test for σ^2

- The power is:

$$\pi = P(\chi^2_\beta < \chi^2 < \infty)$$

where

$$\chi^2_\beta = \chi^2_{1-\alpha}\left(\frac{\sigma_0}{\sigma_1}\right)^2.$$

- The exact sample size is determined by the condition:

$$\frac{\chi^2_{1-\alpha}}{\chi^2_\beta} \leq \left(\frac{\sigma_1}{\sigma_0}\right)^2.$$

df	$\chi^2_{0.95}/\chi^2_{0.10}$	$\chi^2_{0.90}/\chi^2_{0.05}$
2	28.43	44.89
4	8.920	10.95
10	3.763	4.057
20	2.524	2.618

df	$\chi^2_{0.95}/\chi^2_{0.10}$	$\chi^2_{0.90}/\chi^2_{0.05}$
50	1.791	1.817
100	1.510	1.521
200	1.400	1.407
500	1.203	1.204

Test for σ^2

- When the sample size is large, the distribution of $\ln(s)$ is approximately normal with $\mu_{\ln(s)} = \ln(\sigma)$ and $\sigma_{\ln(s)} = 1/\sqrt{2n}$.
- The power by the large sample approximation is:

$$\pi = \Phi(-z_\beta < z < \infty)$$

where

$$z_\beta = \sqrt{2n}\,\ln\left(\frac{\sigma_1}{\sigma_0}\right) - z_\alpha.$$

- The approximate sample size required to obtain power $\pi = 1 - \beta$ is:

$$n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{\ln(\frac{\sigma_1}{\sigma_0})}\right)^2.$$

Test for σ^2

Example: What sample size is required to reject $H_0 : \sigma^2 = 40$ in favor of $H_A : \sigma^2 > 40$ with $\pi = 0.90$ when $\sigma^2 = 100$?

Solution: With $z_{0.05} = 1.645$ and $z_{0.10} = 1.282$:

$$n = \frac{1}{2}\left(\frac{1.645 + 1.282}{\ln\left(\sqrt{\frac{100}{40}}\right)}\right)^2 = 21.$$

The solution by the exact method gives $n = 22$.

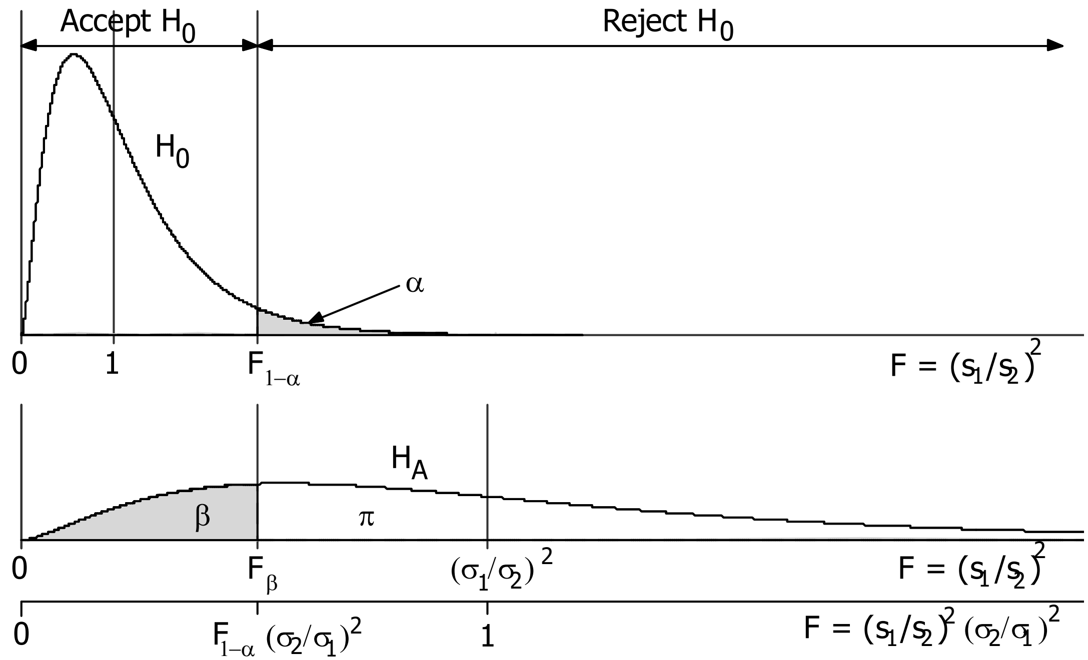
Test for Two Standard Deviations

- The hypotheses to be tested are $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_A : \sigma_1^2 > \sigma_2^2$.
- The test statistic is

$$F = \left(\frac{s_1}{s_2}\right)^2.$$

- The exact power is given by:

$$\pi = P(F_\beta < F < \infty) \text{ where } F_\beta = \left(\frac{\sigma_2}{\sigma_1}\right)^2 F_{1-\alpha}.$$



Test for Two Standard Deviations

- The distribution of $\ln(s_1/s_2)$ is approximately normal.
- The large-sample approximate power is:

$$\pi = \Phi(-z_\beta < z < \infty)$$

where

$$z_\beta = \frac{\ln(\frac{\sigma_1}{\sigma 2})}{\sqrt{\frac{1}{2}(\frac{1}{n_1} + \frac{1}{n_2})}} - z_\alpha.$$

- When $n_1 = n_2$, the approximate sample size is:

$$n_1 = n_2 = \left(\frac{z_\alpha + z_\beta}{\ln(\frac{\sigma_1}{\sigma 2})} \right)^2.$$

Test for Two Standard Deviations

Example: What sample size is required to reject $H_0 : \sigma = 10$ in favor of $H_A : \sigma > 10$ with $\pi = 0.90$ when $\sigma = 20$?

Solution: With $\sigma_1/\sigma_0 = 20/10 = 2$, $z_{0.05} = 1.645$, and $z_{0.10} = 1.282$:

$$n_1 = n_2 = \left(\frac{z_\alpha + z_\beta}{\ln(\frac{\sigma_1}{\sigma 2})} \right)^2$$

$$n_1 = n_2 = \left(\frac{z_{0.05} + z_{0.10}}{\ln(2)} \right)^2$$

$$= 18$$

MINITAB's **Stat**> **Power and Sample Size**> **2 Variances** menu gives $n = 20$.

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. **Proportions**
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Proportions

- One Proportion
  - Exact (binomial) method
  - χ^2 approximation
  - Normal approximations
  - Larson's nomogram
- Two Proportions
  - Difference
  - Risk ratio or relative risk
  - Odds ratio
  - Fisher's exact test
  - McNemar's test for correlated proportions

Proportions

- The distribution of successes (x) in trials (n), when the probability of a success in any trial (p) is fixed, follows the binomial model:

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

- Exact confidence intervals and hypothesis test are performed with the binomial model but there are many approximations available.
- If a large-sample approximation gives a sample size that's large compared to the population size, use the small-population correction

$$n' = \frac{n}{1 + \frac{n-1}{N}}$$

where n is the sample size obtained by the large-sample method and n' is the corrected sample size.

Confidence Interval for One Proportion

- When p is small, the approximate sample size required to demonstrate the one-sided upper confidence limit for the population proportion with the form:

$$P(0 < p < p_U) = 1 - \alpha$$

is given by:

$$n \simeq \frac{\chi^2_{1-\alpha,2(X+1)}}{2p_U}$$

where X is the number of successes in the sample.

- For the special case of $P(0 < p < p_U) = 0.95$, without any successes found in the sample, the approximate sample size is given by the *rule of three*:

$$n \simeq \frac{\chi^2_{0.95,2}}{2p_U} \simeq \frac{3}{p_U}.$$

Confidence Interval for One Proportion

Example: How many units must be inspected without any failures to be 95% confident that the defective rate is less than 1%?

Solution: The desired confidence interval has the form:

$$P(0 < p < 0.01) = 0.95.$$

By the rule of three, the sample size must be:

$$n \simeq \frac{3}{0.01} \simeq 300.$$

Confidence Interval for One Proportion

- When the sample size is large and $0.1 < p < 0.9$, the confidence interval for p is approximately:

$$P(\hat{p} - \delta < p < \hat{p} + \delta) = 1 - \alpha$$

where

$$\delta = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

which leads to

$$n \simeq \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$

- When $p \simeq \frac{1}{2}$ and $\alpha = 0.05$:

$$n \simeq \frac{1}{4}\left(\frac{z_{0.025}}{\delta}\right)^2$$
$$\simeq \delta^{-2}.$$

Confidence Interval for One Proportion

Example: How many people must be polled in a close election to estimate how the election will go with 2% precision and 95% confidence?

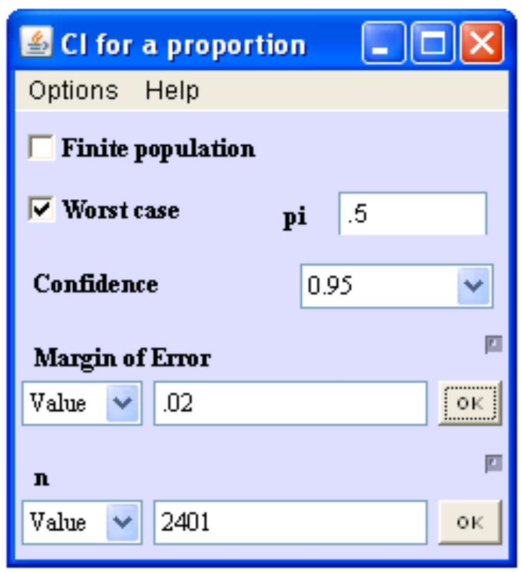
Solution: The desired confidence interval has the form:

$$P(\widehat{p} - 0.02 < p < \widehat{p} + 0.02) = 0.95.$$

The sample size must be:

$$n \simeq (0.02)^{-2} = 2500.$$

Solution using Piface **CI for one proportion**:



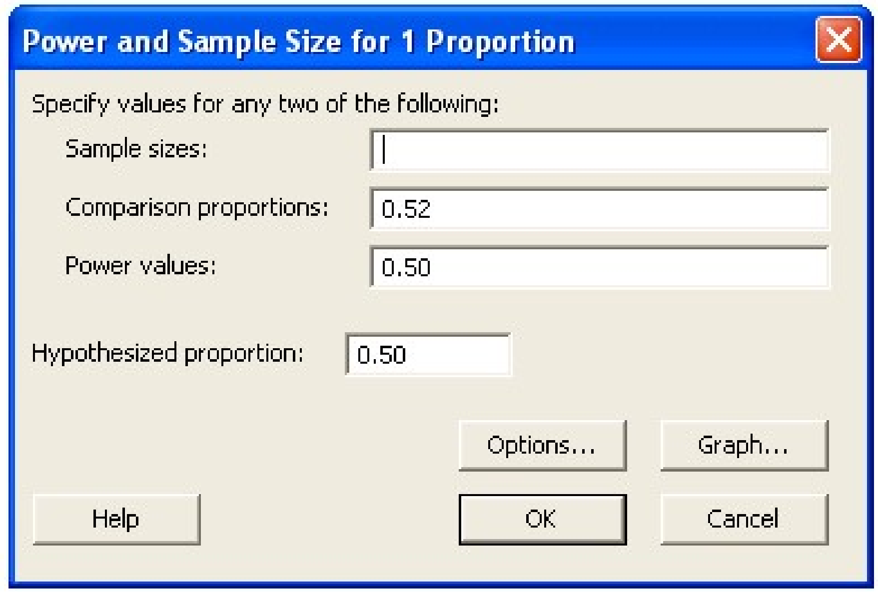
Confidence Interval for One Proportion

Solution using MINITAB **Stat**> **Power and Sample Size**> **1 Proportion**:

```
MTB > Power;
SUBC>   POne;
SUBC>     PCompare 0.52;
SUBC>     Power 0.50;
SUBC>     PNull 0.50;
SUBC>   GPCurve.
```

Power and Sample Size

```
Test for One Proportion

Testing p = 0.5 (versus not = 0.5)
Alpha = 0.05


                Sample  Target
Comparison p     Size   Power   Actual Power
        0.52     2401     0.5       0.500058
```



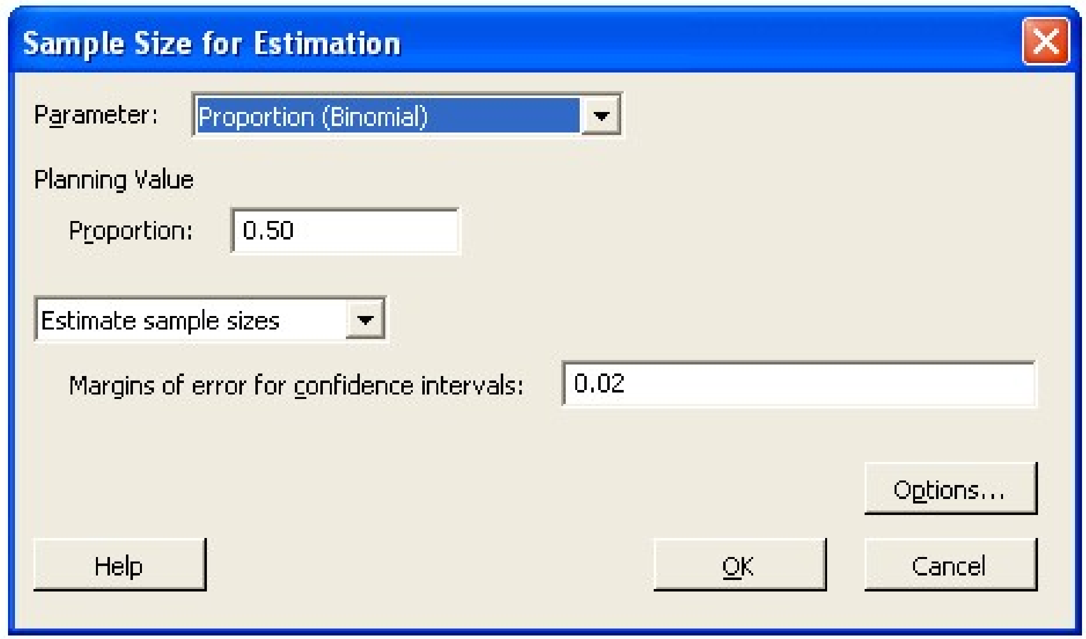
Confidence Interval for One Proportion

Solution using MINITAB **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Proportion** (**Binomial**):

```
MTB > SSCI;
SUBC>   BProportion 0.50;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 0.02.
```

Sample Size for Estimation

```
Method

Parameter            Proportion
Distribution         Binomial
Proportion           0.5
Confidence level     95%
Confidence interval  Two-sided


Results

  Margin   Sample
of Error     Size
    0.02     2449
```



Proportions In Small Populations

- When the population is small, attention shifts from the success rate to the number of successes.
- The hypergeometric distribution governs attribute sampling from small populations, but there are good approximations to it in most cases.
  - The small-sample ($n \ll N$) binomial approximation:

$$\begin{aligned} h(x; S, N, n) &\simeq b(x; n, p = S/N) \\ &\simeq \binom{n}{x}\left(\frac{S}{N}\right)^{x}\left(1 - \frac{S}{N}\right)^{n-x}. \end{aligned}$$

  - The rare-event ($S \ll N$) binomial approximation:

$$\begin{aligned} h(x; S, N, n) &\simeq b(x; S, p = n/N) \\ &\simeq \binom{S}{x}\left(\frac{n}{N}\right)^{x}\left(1 - \frac{n}{N}\right)^{S-x}. \end{aligned}$$

Confidence Interval for One Proportion (Small Population)

- The one-sided upper $(1-\alpha)100\%$ confidence interval for S is given by

$$P(S \le S_U) \ge 1-\alpha$$

where S_U is the smallest value of S which satisfies

$$\sum_{x=0}^{X} h(x; S_U, N, n) \le \alpha.$$

Confidence Interval for One Proportion (Small Population)

- When $X = 0$ and $n \ll N$:

$$h(0; S_U, N, n) \simeq b\left(0; n, p = \frac{S_U}{N}\right) \simeq \left(1 - \frac{S_U}{N}\right)^n$$

which leads to:

$$n \simeq \frac{\ln(\alpha)}{\ln\left(1 - \frac{S_U}{N}\right)}.$$

This result is equivalent to:

$$n \simeq \frac{\chi^2_{1-\alpha,2}}{2\left(\frac{S_U}{N}\right)}.$$

Confidence Interval for One Proportion (Small Population)

- When $X = 0$ and $S \ll N$:

$$h(0; S_U, N, n) \simeq b\left(0; S_U, p = \frac{n}{N}\right)$$
$$\simeq \left(1 - \frac{n}{N}\right)^{S_U}$$

which leads to:

$$n \geq N(1 - \alpha^{1/S_U})$$

or

$$\frac{n}{N} \geq 1 - \alpha^{1/S_U}.$$

Confidence Interval for One Proportion (Small Population)

Example: What fraction of a lot must be inspected and found to be free of defectives to demonstrate,with 95% confidence, that there are no more than four defectives in the population?

Solution: The goal of the experiment is to demonstrate the confidence interval

$$P(0 \leq S \leq 4) \geq 0.95$$

using a zero-successes $(X = 0)$ sampling plan. By the rare-event approximation,

$$\begin{aligned} \frac{n}{N} &\geq 1 - \alpha^{1/S_U} \\ &\geq 1 - 0.05^{1/4} \\ &\geq 0.53. \end{aligned}$$

Test for One Proportion

- The hypotheses to be tested are $H_0 : p = p_0$ versus $H_A : p > p_0$.
- The exact power and sample size are determined by the simultaneous solution to

$$\sum_{x=0}^{c} b(x; n, p_0) \geq 1 - \alpha$$

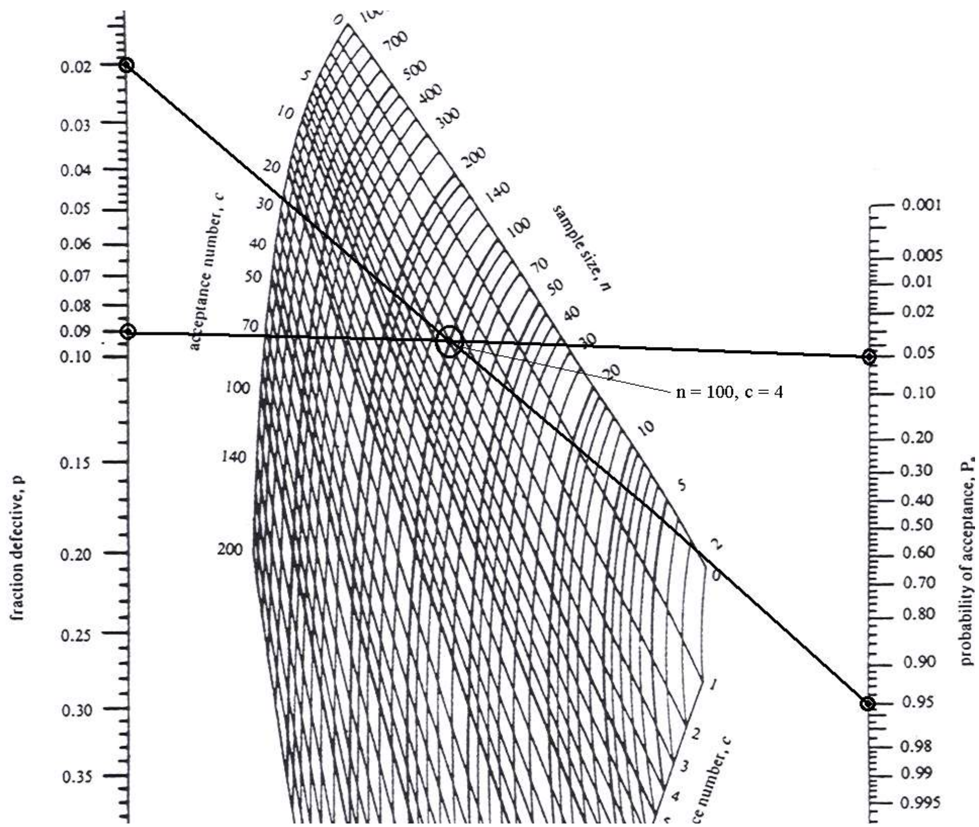
$$\sum_{x=0}^{c} b(x; n, p_1) \leq \beta.$$

  where c is the acceptance number.
- Use a large-sample approximation to find an approximate solution, then iterate the equations to find an exact solution.
- Larson's nomogram is the easiest way to find an approximate solution.

Test for One Proportion

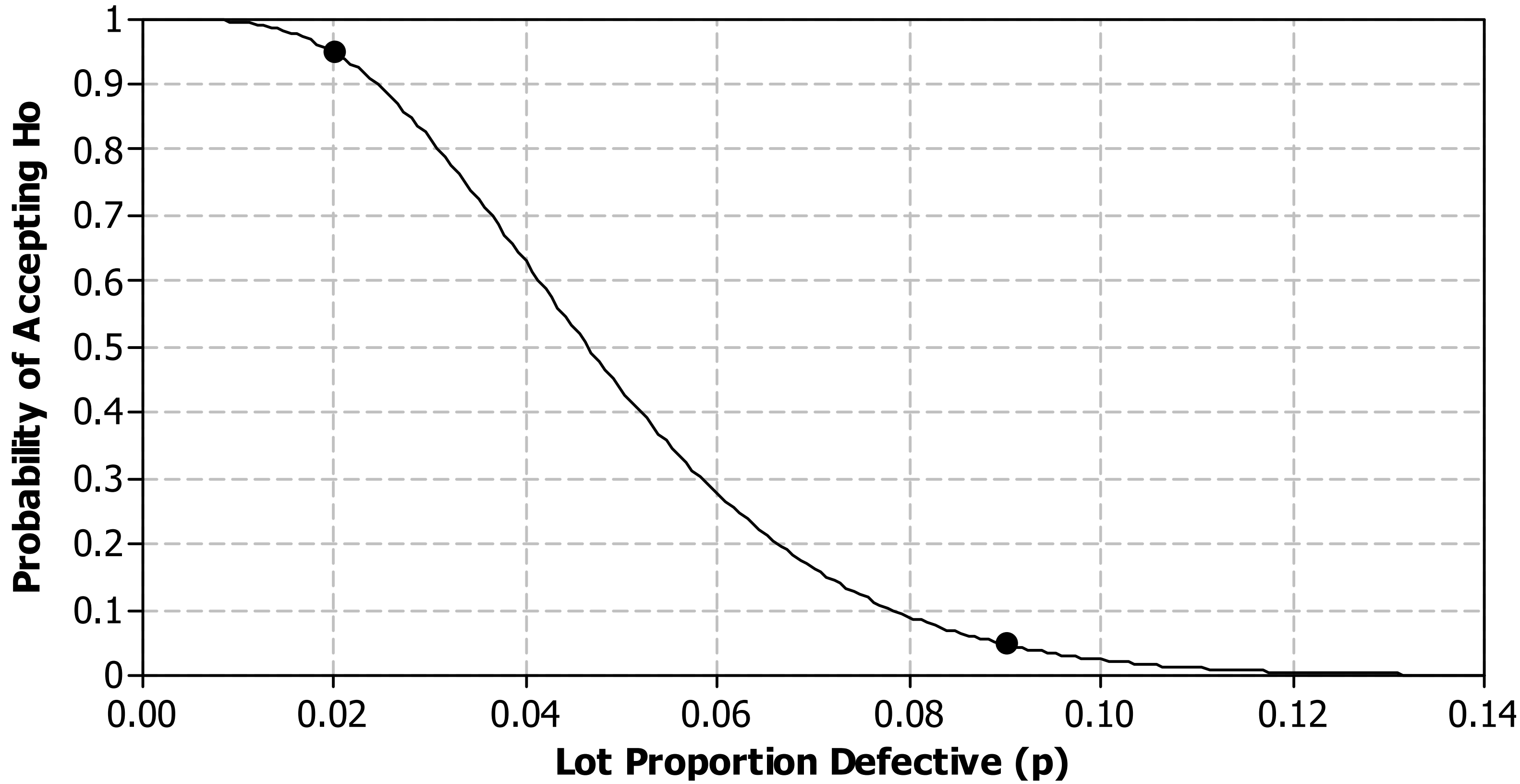
Example: Find n and c for the sampling plan for defectives that will accept 95% of lots with 2% defectives and 5% of lots with 9% defectives. Draw the OC curve for the sampling plan.

Solution:



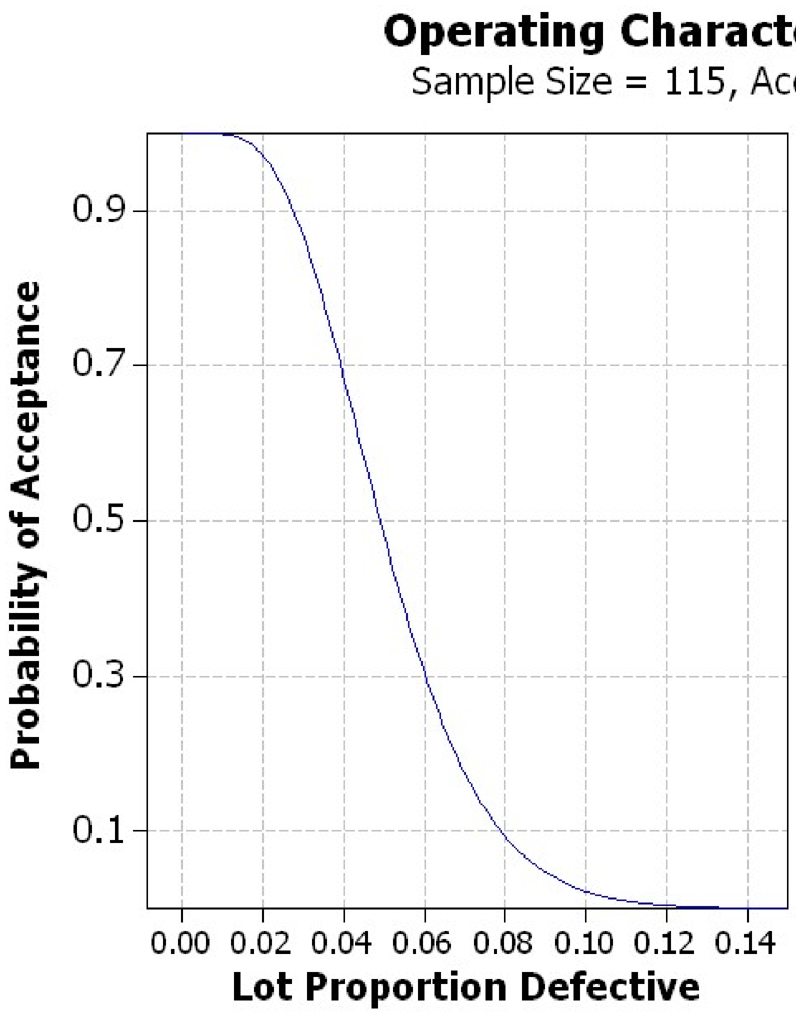
Test for One Proportion

Solution (continued):



Test for One Proportion

Solution by MINITAB **Stat**> **Quality Tools**> **Acceptance Sampling by Attributes**> **Create a Sampling Plan**:




```
Acceptance Sampling by Attributes

Measurement type:  Go/no go
Lot quality in proportion defective
Use binomial distribution to calculate probability
of acceptance


Acceptable Quality Level (AQL)          0.02
Producer's Risk (Alpha)                 0.05

Rejectable Quality Level (RQL or LTPD)  0.09
Consumer's Risk (Beta)                  0.05


Generated Plan(s)

Sample Size        115
Acceptance Number    5

Accept lot if defective items in 115 sampled <= 5,
otherwise reject.


Proportion  Probability  Probability
 Defective    Accepting    Rejecting
      0.02        0.971        0.029
      0.09        0.047        0.953
```

Confidence Interval for the Difference Between Two Proportions

- The two-sided confidence interval for Δp has the form

$$P(\Delta\widehat{p} - \delta < \Delta p < \Delta\widehat{p} + \delta) = 1 - \alpha$$

where the confidence interval half-width is

$$\delta = z_{\alpha/2}\widehat{\sigma}_{\Delta\widehat{p}}.$$

- Then the sample size n_1 required to obtain the desired confidence interval half-width δ with sample size ratio n_1/n_2 is

$$n_1 = \left(\frac{z_{\alpha/2}}{\delta}\right)^2 \left(p_1(1-p_1) + p_2(1-p_2)\left(\frac{n_1}{n_2}\right)\right).$$

- If p_1 and p_2 are expected to be approximately equal so that they can both be estimated by a nominal value p, then

$$n_1 = \left(\frac{z_{\alpha/2}}{\delta}\right)^2 \left(1 + \frac{n_1}{n_2}\right)p(1-p).$$

Test for a Difference Between Two Proportions

- The hypotheses to be tested are $H_0 : p_1 = p_2$ versus $H_A : p_1 \neq p_2$.
- The power of the test is

$$\pi = \Phi(-\infty < z < z_\beta)$$

where

$$z_\beta = \frac{\Delta\hat{p}}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}} - z_{\alpha/2}$$

and

$$\hat{p} = \frac{1}{2}(\hat{p}_1 + \hat{p}_2).$$

- The sample size is

$$n = \frac{2\hat{p}(1-\hat{p})}{(\Delta\hat{p})^2}(z_{\alpha/2} + z_\beta)^2.$$

Test for a Difference Between Two Proportions

Example: A biologist wants to test for a difference in the ratio of male to female frogs between a clean pond and a contaminated pond. How many frogs must she sample to detect a difference of 10% between the ponds with 90% power?

Solution: Assuming that normal ratio of male to female frogs is $1 : 1$, with $\widehat{p} = 0.5$ and $\Delta\widehat{p} = 0.10$, the approximate sample size is:

$$n = \frac{2(0.5)(1-0.5)}{(0.10)^2}(1.96 + 1.282)^2 = 526.$$

Test for a Difference Between Two Proportions

Solution by Piface **Test Comparing Two Proportions**:

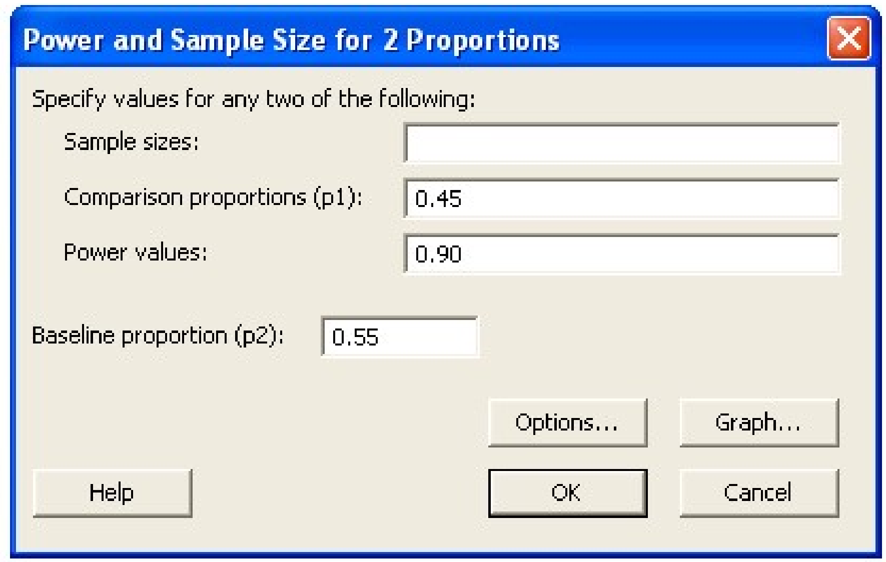
Test for a Difference Between Two Proportions

Solution by MINITAB **Stat**> **Power and Sample Size**> **2 Proportions**:

```
MTB > Power;
SUBC>   PTwo;
SUBC>     PCompare 0.45;
SUBC>     Power 0.90;
SUBC>     PBaseline 0.55;
SUBC>   GPCurve.
```

Power and Sample Size

```
Test for Two Proportions

Testing comparison p = baseline p (versus not =)
Calculating power for baseline p = 0.55
Alpha = 0.05


              Sample  Target
Comparison p    Size   Power  Actual Power
        0.45     524     0.9      0.900386

The sample size is for each group.
```



Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. **Counts**
    a. **One population**
    b. **Two populations**
    c. **Many populations**
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Poisson Counts

- The distribution of counts is Poisson:

$$Poisson(x; n, \lambda) = \frac{(n\lambda)^x e^{-n\lambda}}{x!} \text{ for } x = 0, 1, \ldots$$

  where x is the number of counts observed in area of opportunity n. The mean of x is $\mu_x = n\lambda$, where λ is the mean x per unit area.
- When the Poisson mean is large, the Poisson distribution is approximately normal.
- The distribution of $\sqrt{x}$ is approximately normal with mean $\mu_{\sqrt{x}} = \sqrt{n\lambda}$ and standard deviation $\sigma_{\sqrt{x}} = \frac{1}{2}$. This transformation provides convenient methods for sample size and power calculations.

Confidence Interval for the Poisson Mean

- An approximate large sample confidence interval for the Poisson mean is

$$P\left(\frac{x - z_{\alpha/2}\sqrt{x}}{n} < \lambda < \frac{x + z_{\alpha/2}\sqrt{x}}{n} \right) = 1 - \alpha$$

$$P\left(\hat{\lambda}(1 - \delta) < \lambda < \hat{\lambda}(1 + \delta) \right) = 1 - \alpha$$

where

$$\hat{\lambda} = \frac{x}{n} \text{ and } \delta = z_{\alpha/2}/\sqrt{x}.$$

- The number of events x required to obtain a specified value of δ is given by

$$x = \left(\frac{z_{\alpha/2}}{\delta} \right)^2.$$

Confidence Interval for the Poisson Mean

Example: How much junk mail must be accumulated to estimate the daily rate of junk mail with 10% precision and 95% confidence?

Solution: The desired confidence interval has the form

$$P\left(\hat{\lambda}(1 - 0.10) < \lambda < \hat{\lambda}(1 + 0.10)\right) = 0.95.$$

The total number of pieces of junk mail required is

$$\begin{aligned} x &= \left(\frac{z_{0.025}}{0.1}\right)^2 \\ &= \left(\frac{1.96}{0.1}\right)^2 \\ &= 385. \end{aligned}$$

After at least that many pieces of junk mail are collected,

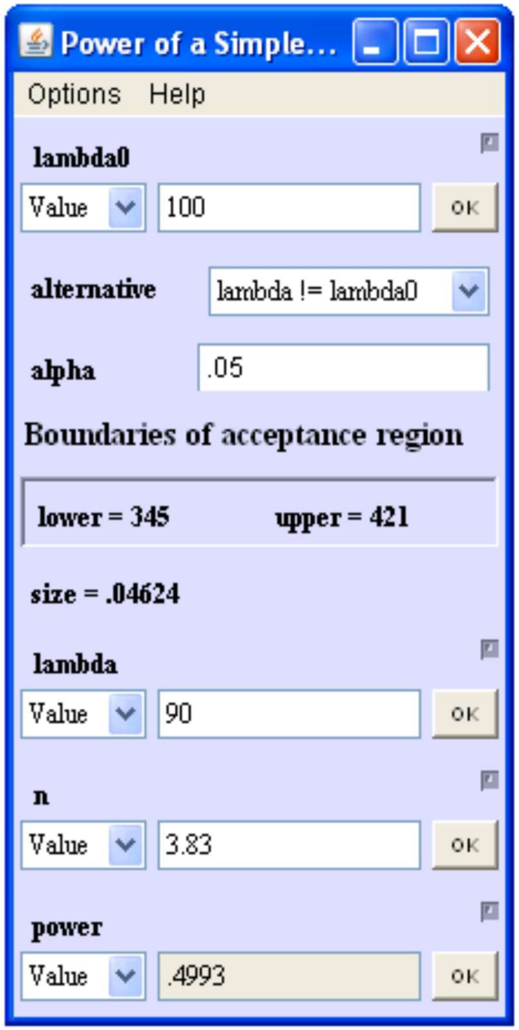
$$\hat{\lambda} = \frac{x}{n}$$

where n is the number of days, and the confidence limits for λ are

$$UCL/LCL = (1 \pm 0.1)\hat{\lambda}.$$

Confidence Interval for the Poisson Mean

Solution: By Piface, $x = 100 \times 3.83 = 383$:



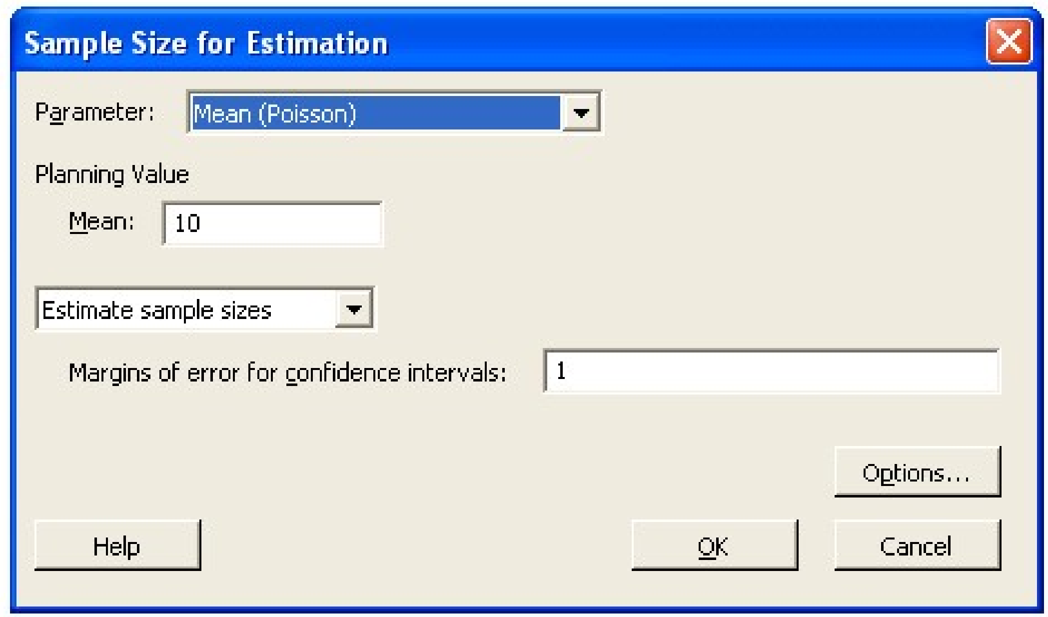
Confidence Interval for the Poisson Mean

Solution using MINITAB **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Mean Poisson**:

```
MTB > SSCI;
SUBC>   PMean 10;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 1.
```

Sample Size for Estimation

```
Method

Parameter              Mean
Distribution           Poisson
Mean                   10
Confidence level       95%
Confidence interval    Two-sided


Results

  Margin   Sample
of Error     Size
       1       43
```




```
So the total number of counts that must be accumulated is 43 x 10 = 430.
```

Test for One Poisson Mean

- The hypotheses to be tested are $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda > \lambda_0$.
- The approximate power is

$$\pi = \Phi(-z_\beta < z < \infty)$$

where

$$z_\beta = -2\sqrt{n}\left(\sqrt{\lambda_1} - \sqrt{\lambda_0}\right) + z_\alpha.$$

- The number of units that must be inspected to reject H_0 with specified power is

$$n = \frac{1}{4}\left(\frac{z_\alpha + z_\beta}{\sqrt{\lambda_1} - \sqrt{\lambda_0}}\right)^2.$$

Test for One Poisson Mean

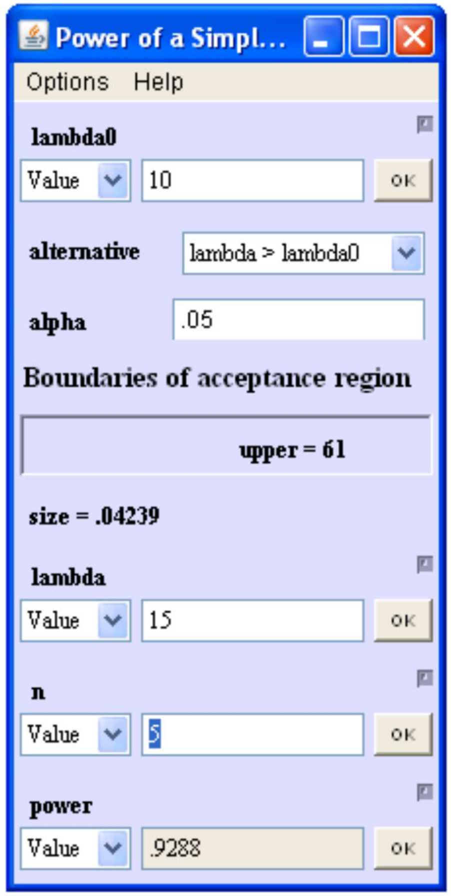
Example: How many sampling units must be inspected to reject $H_0 : \lambda = 10$ with 90% power in favor of $H_A : \lambda > 10$ when $\lambda = 15$?

Solution:

$$
\begin{aligned}
n &= \frac{1}{4}\left(\frac{z_{0.05} + z_{0.10}}{\sqrt{15} - \sqrt{10}}\right)^2 \\
&= \frac{1}{4}\left(\frac{1.645 + 1.282}{\sqrt{15} - \sqrt{10}}\right)^2 \\
&= 4.24
\end{aligned}
$$

Test for One Poisson Mean

Using Piface **Generic Poisson Test** the sample size must be $n = 5$.



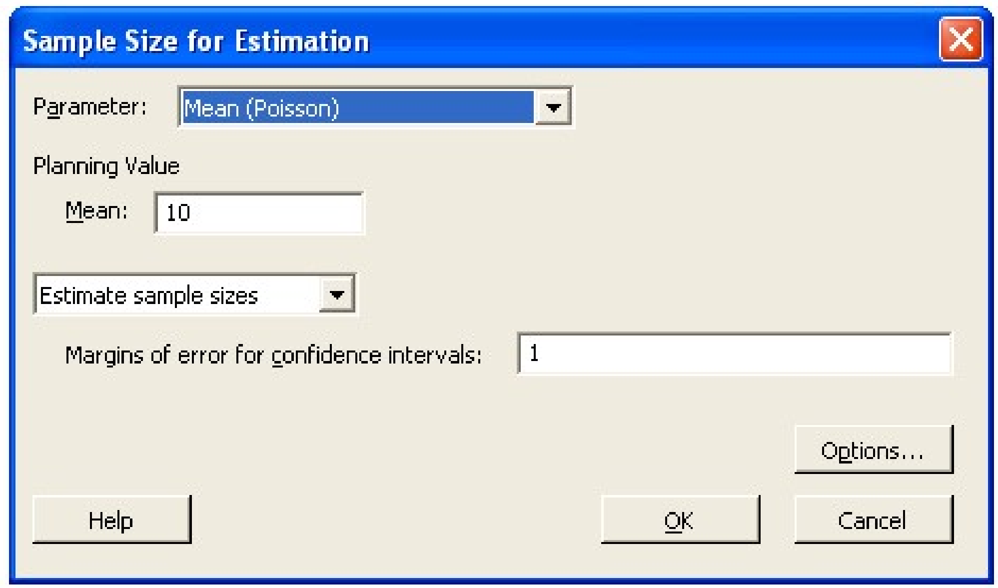
Test for One Poisson Mean

Solution using MINITAB **Stat**> **Power and Sample Size**> **1-Sample Poisson Rate**:

```
MTB > Power;
SUBC>   OneRate;
SUBC>   RCompare 15;
SUBC>   Power 0.90;
SUBC>   RNull 10;
SUBC>   Alternative 1;
SUBC>   Alpha 0.05;
SUBC>   Length 1.0;
SUBC>   GPCurve.
```

Power and Sample Size

```
Test for 1-Sample Poisson Rate

Testing rate = 10 (versus > 10)
Alpha = 0.05
"Length" of observation = 1


Comparison  Sample  Target
      Rate    Size   Power  Actual Power
        15       5     0.9      0.938674
```



Test for Two Poisson Means

- The hypotheses to be tested are $H_0 : \lambda_1 = \lambda_2$ versus $H_A : \lambda_1 < \lambda_2$.
- The approximate power is

$$\pi = \Phi(-\infty < z < z_\beta)$$

where

$$n_1 = \frac{1}{4}\left(1 + \frac{n_1}{n_2}\right)\left(\frac{z_\alpha + z_\beta}{\sqrt{\lambda_2} - \sqrt{\lambda_1}}\right)^2$$

and n_1/n_2 is the sample size allocation ratio.

- The optimum allocation ratio is

$$\frac{n_1}{n_2} = \sqrt{\frac{\lambda_1}{\lambda_2}}.$$

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. **Linear Regression**
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Linear Regression

- The linear regression model is

$$y_i = b_0 + b_1 x_i + \epsilon_i$$

- The usual goal in linear regression is to estimate the slope, e.g.:

$$P(b_1 - \delta < \beta_1 < b_1 + \delta) = 1 - \alpha$$

where

$$\begin{aligned} \delta &= t_{\alpha/2} \widehat{\sigma}_{b_1} \\ &= t_{\alpha/2} \left(\frac{\widehat{\sigma}_\epsilon}{\sqrt{SS_x}} \right) \end{aligned}$$

and

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

- The confidence interval half-width depends on the pattern of x values.

Linear Regression

- If x is normal:

$$N \geq \left(\frac{t_{\alpha/2} \widehat{\sigma}_{\epsilon}}{\delta \widehat{\sigma}_x} \right)^2$$

- k evenly spaced, equally weighted levels of x:

$$N \geq \frac{12}{(k-1)(k+1)} \left(\frac{t_{\alpha/2} \widehat{\sigma}_{\epsilon}}{\delta \Delta x} \right)^2 \text{ where } \Delta x = \frac{x_{max} - x_{min}}{k-1}$$

- If x is uniformly distributed between $x_{\min}$ and $x_{\max}$:

$$N \geq 12 \left(\frac{t_{\alpha/2} \widehat{\sigma}_{\epsilon}}{\delta (x_{max} - x_{min})} \right)^2$$

- $k = 2$ levels of x:

$$N \geq 4 \left(\frac{t_{\alpha/2} \widehat{\sigma}_{\epsilon}}{\delta (x_{max} - x_{min})} \right)^2$$

Linear Regression

- Comparison of the total number of observations under different patterns for the x observations to obtain the same estimation precision for the slope:
  - Three levels of x versus two levels of x:

$$\frac{N_{\text{three levels of } x}}{N_{\text{two levels of } x}} = 1.5$$

  - Uniform distribution of x versus two levels of x:

$$\frac{N_{\text{uniform distribution of } x}}{N_{\text{two levels of } x}} = 3$$

- Conclusions:
  - Pick the range of x to be as wide as is practically possible.
  - Concentrate observations at the x extremes
  - If a lack of fit test is required, add observations at the center of the x range.

Test for the Regression Slope

- The hypotheses to be tested are $H_0 : \beta_1 = 0$ versus $H_A : \beta_1 \neq 0$, where β_1 is the regression slope parameter.
- The power to reject H_0 is

$$\pi = P(-\infty < t < t_\beta)$$

where

$$t_\beta = \frac{|\beta_1| \sqrt{SS_x}}{\sigma_\epsilon} - t_{\alpha/2}$$

and

$$SS_x = N\sigma_x^2.$$

- The number of observations required to obtain a specified power value for a given β_1 value is the smallest value of N that satisfies

$$N \geq (t_{\alpha/2} + t_\beta)^2 \left(\frac{\sigma_\epsilon}{\beta_1 \sigma_x} \right)^2.$$

Test for the Regression Slope

Example: How many observations are required to reject $H_0 : \beta_1 = 0$ in favor of $H_A : \beta_1 \neq 0$ with 90% power for $\beta_1 = 10$ when $\sigma_x \simeq 2$ and $\sigma_\epsilon = 30$?

Solution: With $t \simeq z$, the first iteration gives

$$N = (z_{0.025} + z_{0.10})^2 \left(\frac{30}{10 \times 2} \right)^2 = 24.$$

Further iterations indicate that the required sample size is $N = 26$.

Using Piface **Linear Regression**:

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. **Correlation**
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

Correlation

- Pearson's correlation coefficient is ρ, where $\rho = 0$ indicates no correlation, $\rho = 1$ indicates perfect positive correlation, and $\rho = -1$ indicates perfect negative correlation.
- The distribution of Fisher's Z transform given by

$$\begin{aligned} Z &= \tanh^{-1}(r) \\ &= \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \end{aligned}$$

is approximately normal with mean

$$\mu_Z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

and standard deviation

$$\sigma_Z = \frac{1}{\sqrt{n-3}}.$$

Confidence Interval for the Correlation Coefficient

- If numerical values are chosen for the upper and lower confidence limits of ρ,

$$P(LCL_\rho < \rho < UCL_\rho) = 1 - \alpha,$$

then the limits may be Z-transformed to obtain

$$P(Z_{LCL_\rho} < Z_\rho < Z_{UCL_\rho}) = 1 - \alpha.$$

- The sample size to obtain the desired confidence interval is

$$n = 4\left(\frac{z_{\alpha/2}}{Z_{UCL_\rho} - Z_{LCL_\rho}}\right)^2 + 3.$$

Confidence Interval for the Correlation Coefficient

Example: Determine the number of paired observations required to obtain the following confidence interval for the population correlation:

$$P(0.9 < \rho < 0.99) = 0.95.$$

Solution: The Fisher's Z-transformed confidence interval is

$$P(Z_{0.9} < Z_\rho < Z_{0.99}) = 0.95$$

$$P(1.472 < Z_\rho < 2.647) = 0.95.$$

The required sample size is

$$\begin{aligned} n &= 4\left(\frac{1.96}{2.647 - 1.472}\right)^2 + 3 \\ &= 15. \end{aligned}$$

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. **Designed Experiments**
   a. **One-way ANOVA**
   b. **Balanced Full Factorial Design with Fixed Effects**
   c. **Fixed Effects in Mixed Models**
   d. **Random Effects in Mixed Models**
   e. **Two-level Factorial Designs**
9. Reliability
10. Statistical Quality Control
11. Resampling Methods

One-way ANOVA

- The hypotheses to be tested are $H_0 : \mu_i = \mu_j$ *for all pairs of k treatments* versus $H_A : \mu_i \neq \mu_j$ *for at least one pair of treatments*.
- The test is performed using the F statistic

$$F = \frac{ns_{\bar{x}}^2}{s_\epsilon^2}$$

where the F distribution has $\nu_1 = k - 1$ and $\nu_2 = k(n-1)$ degrees of freedom.

- The acceptance interval for H_0 is

$$P(0 < F < F_{1-\alpha}) = 1 - \alpha.$$

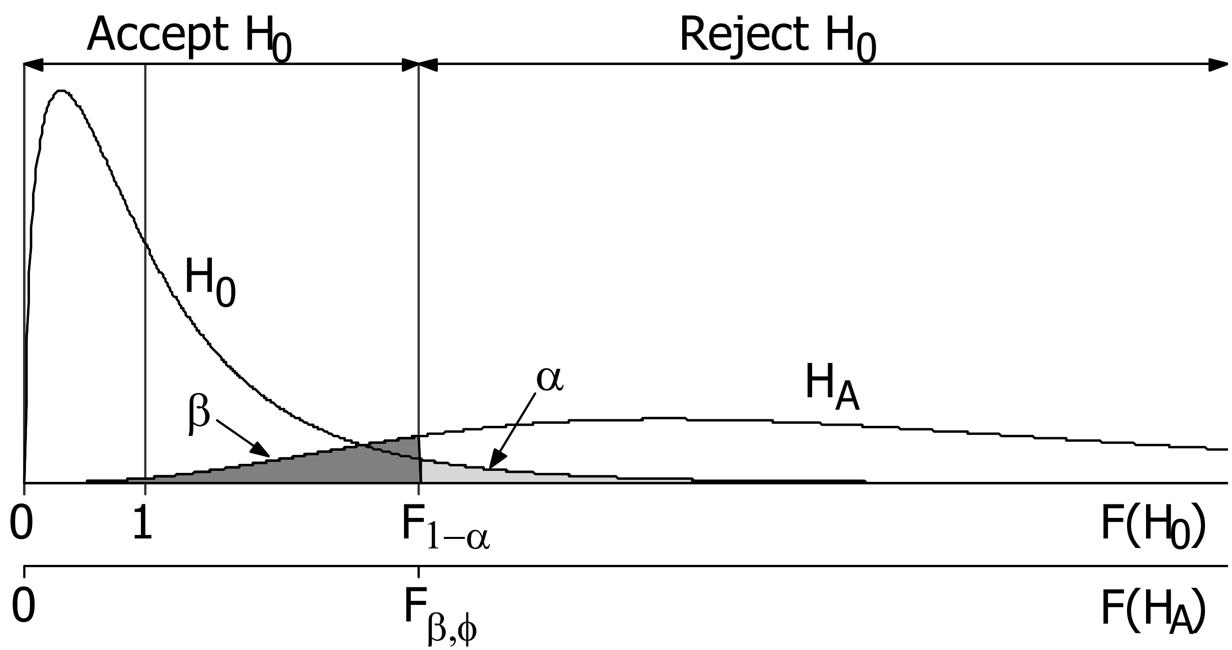
One-way ANOVA

- The power to reject H_0 is given by the condition

$$F_{1-\alpha} = F_{\beta,\phi}$$

where the noncentrality parameter is

$$\phi = \frac{E(SS_{treatments})}{E(MS_{\epsilon})} = \frac{n\sum_{i=1}^{k}\tau_i^2}{\sigma_\epsilon^2}.$$



One-way ANOVA

- If two treatments are biased symmetrically about the mean:

$$\tau_i = \left\{ -\frac{\delta}{2}, \frac{\delta}{2}, 0, 0, \cdots \right\},$$

  then

$$\phi = \frac{n}{2} \left(\frac{\delta}{\sigma_\epsilon} \right)^2.$$

- If one treatment is biased relative to the others:

$$\tau_i = \left\{ -\frac{(k-1)\delta}{k}, \frac{\delta}{k}, \frac{\delta}{k}, \cdots \right\},$$

  then

$$\phi = \frac{n(k-1)}{k} \left(\frac{\delta}{\sigma_\epsilon} \right)^2.$$

- The first condition (two treatments biased symmetrically about the mean) has a smaller noncentrality parameter, i.e. is harder to detect, than the second condition (one treatment biased with respect to all of the others) so the first condition is the one that's usually assumed.

ANOVA Power and Sample Size Calculations

- The condition

$$F_{1-\alpha} = F_{\beta,\phi}$$

  can be used to calculate the power for a specified sample size; however, it cannot be solved explicitly for the sample size as a function of the power. Sample sizes must be determined by iteration.
- An approximate sample size for ANOVA can be calculated by applying Bonferroni's correction to two-sample *t* tests. This sample size is conservative but it provides a good starting point for iterations to determine the exact sample size.

One-way ANOVA Power

Example: In a one-way classification design with four treatments and five observations per treatment, determine the power of the ANOVA to reject H_0 if the treatment biases from the grand mean are $\tau_i = \{18, -6, -6, -6\}$. The four populations are expected to be normal and homoscedastic with $\sigma_\epsilon = 8$.

Solution: With $\delta = 24$ in the noncentrality parameter equation for one treatment biased relative to the others

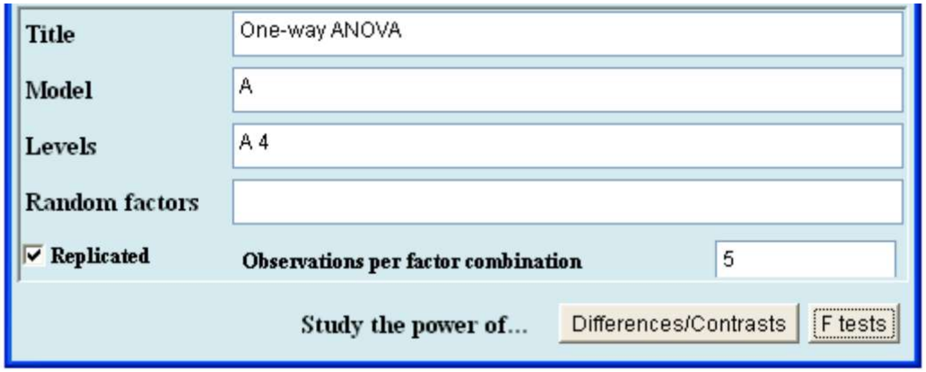
$$\phi = \frac{n(k-1)}{k}\left(\frac{\delta}{\sigma_\epsilon}\right)^2 = \frac{5 \times 3}{4}\left(\frac{24}{8}\right)^2 = 33.75.$$

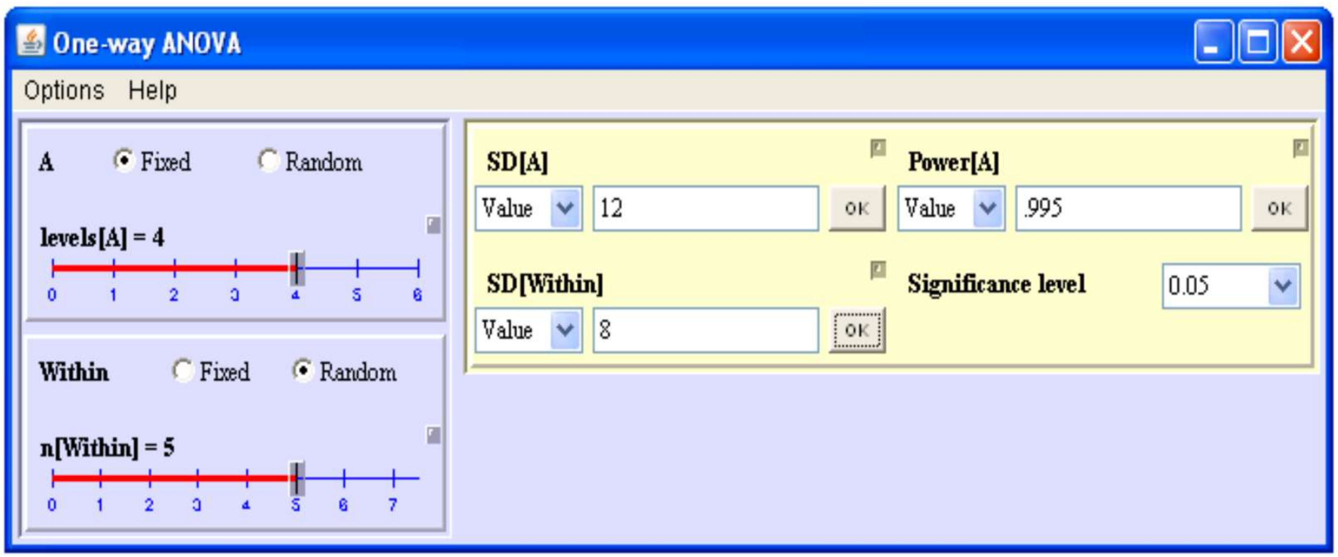
The F statistic will have $df_{treatments} = 4 - 1 = 3$ and $df_\epsilon = 4(5-1) = 16$ degrees of freedom. The power is 99.5% as determined from

$$F_{0.95} = 3.239 = F_{0.005,33.75}.$$

One-way ANOVA

Solution: With $s_A = \sqrt{\left((18)^2 + 3(-6)^2\right)/(4-1)} = 12.0$:



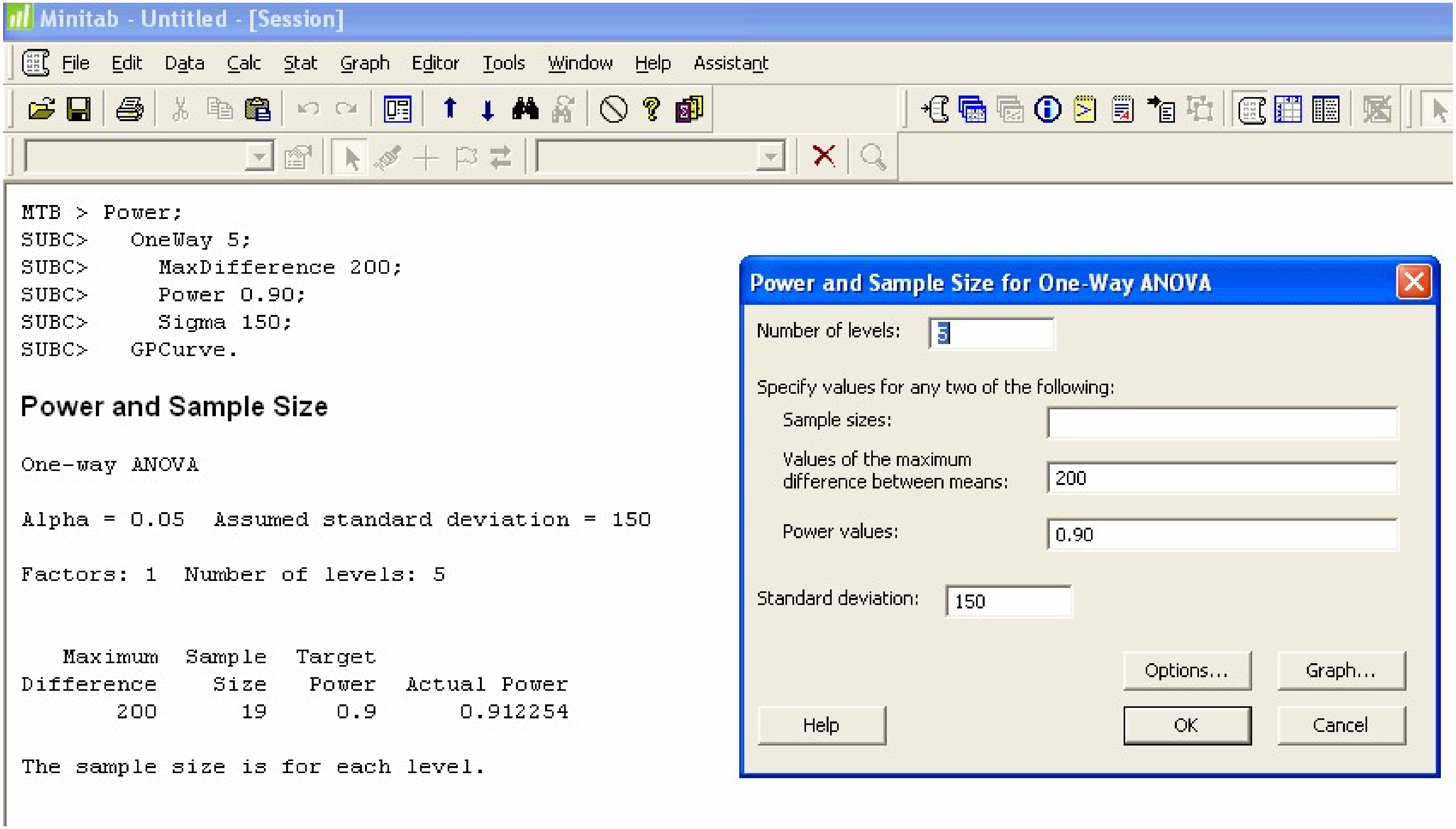


One-way ANOVA

Example: Use MINITAB to determine the sample size required for a one-way classification design with five treatments to be analyzed by ANOVA. The experiment must resolve a difference of $\delta = 200$ with 90% power. The five populations are expected to be normal and homoscedastic with $\sigma_{\epsilon} = 150$. Confirm the value of the power for that sample size.

One-way ANOVA

Solution: From **Stat**> **Power and Sample Size**> **One-Way ANOVA**:



One-way ANOVA

Solution: MINITAB indicates that the experiment requires $n = 19$ observations per treatment group. The model degrees of freedom will be $df_{model} = 5 - 1 = 4$ and the error degrees of freedom will be $df_{\epsilon} = 5(19 - 1) = 90$. The noncentrality parameter (assuming two treatments biased symmetrically about the mean) is

$$\phi = \frac{n}{2}\left(\frac{\delta}{\sigma_{\epsilon}}\right)^2 = \frac{19}{2}\left(\frac{200}{150}\right)^2 = 16.89$$

Then we have:

$$F_{1-\alpha} = F_{\beta,\phi}$$

$$F_{0.95} = 2.486 = F_{\beta,16.89}$$

which is satisfied by $\beta = 0.0877$ so the power is $\pi = 0.9123$ or 91.2%.

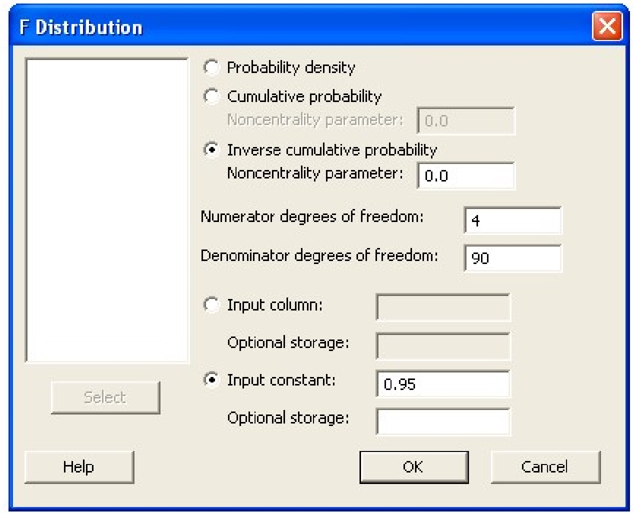
One-way ANOVA

Solution (continued): Using **Calc**> **Probability Distributions**> **F**:

```
MTB > invcdf 0.95;
SUBC> f 4 90.
```

Inverse Cumulative Distribution Function

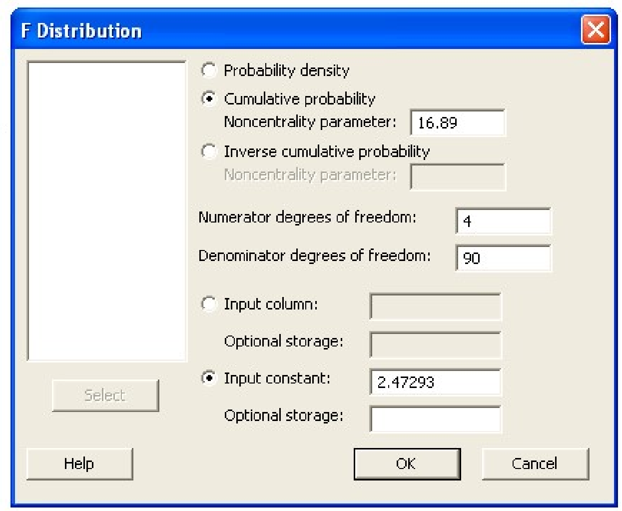
```
F distribution with 4 DF in numerator and 90 DF in denominator

P( X <= x )        x
       0.95  2.47293
```




```
MTB > CDF 2.47293;
SUBC>   F 4 90 16.89.
```

Cumulative Distribution Function

```
F distribution with 4 DF in numerator and 90 DF in denominator and noncentrality parameter 16.89

      x  P( X <= x )
2.47293    0.0877266
```



Balanced Full Factorial Design with Fixed Effects

Example: A $2 \times 3 \times 5$ full factorial experiment with four replicates is planned. The experiment will be blocked on replicates and the ANOVA model will include main effects and two-factor interactions. Determine the power to detect a difference $\delta = 300$ units between two levels of the third study variable if the standard error of the model is expected to be $\sigma_\epsilon = 500$.

Balanced Full Factorial Design with Fixed Effects

Solution: If the three study variables are given the names A, B, and C and have $a = 2$, $b = 3$, and $c = 5$ levels, respectively, then $df_{blocks} = 3$, $df_A = 1$, $df_B = 2$, $df_C = 4$, $df_{AB} = 2$, $df_{AC} = 4$, $df_{BC} = 8$, and $df_\epsilon = 95$. The F distribution noncentrality parameter for C with biases $\delta_1 = -150$, $\delta_2 = 150$, and $\delta_3 = \delta_4 = \delta_5 = 0$ is

$$\begin{aligned}
\phi_C &= \frac{abn \sum_{i=1}^{5} \delta_i^2}{\sigma_\epsilon^2} \\
&= \frac{2 \times 3 \times 4 \times \left((-150)^2 + 150^2 + 0^2 + 0^2 + 0^2\right)}{500^2} \\
&= 4.32.
\end{aligned}$$

The power is determined from

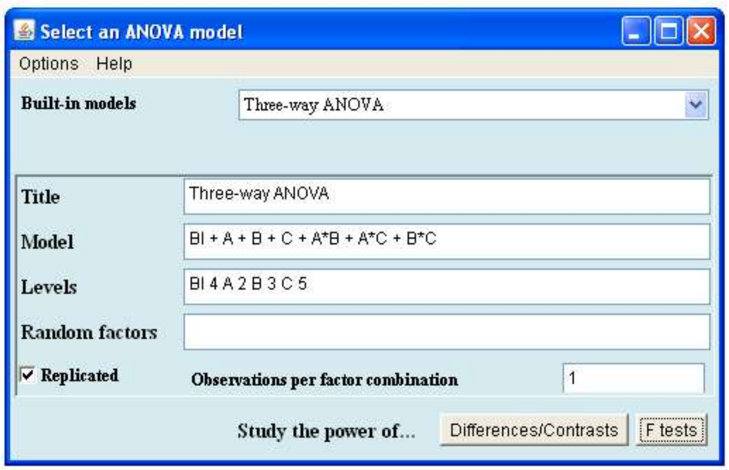
$$F_{0.95} = 2.469 = F_{1-\pi_C, 4.32},$$

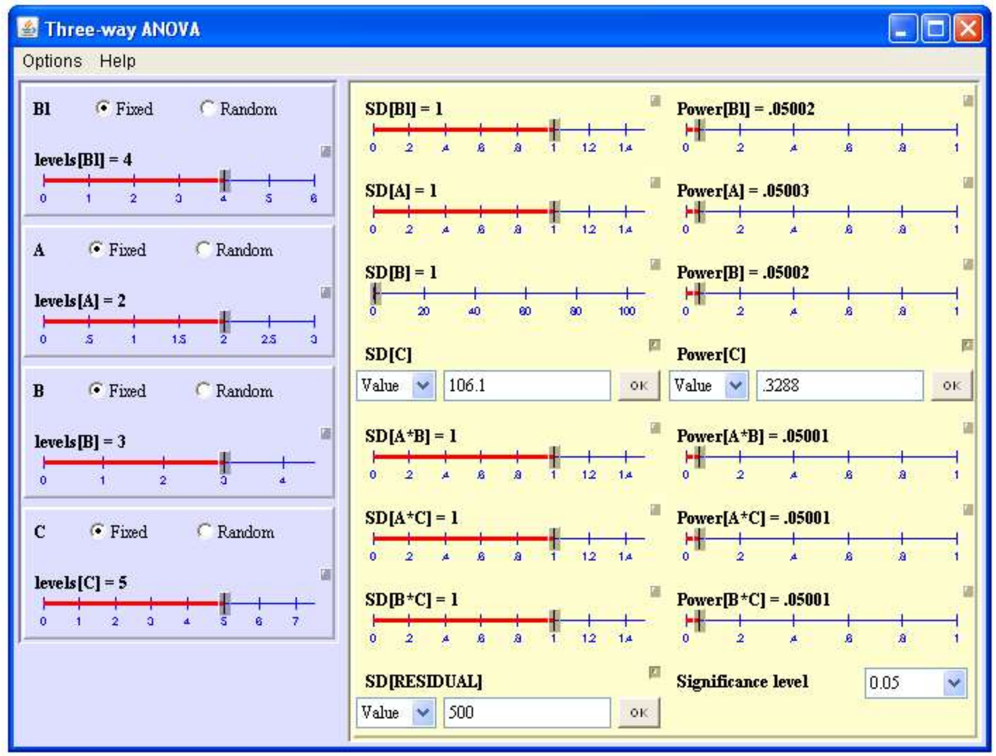
which is satisfied by $\pi_C = 0.328$.

Balanced Full Factorial Design with Fixed Effects

Solution: Using Piface **Balanced ANOVA**> **Three-way Design** with

$$s_C = \sqrt{\left(2(150)^2 + 3(0)^2\right)/(5-1)} = 106.1:$$





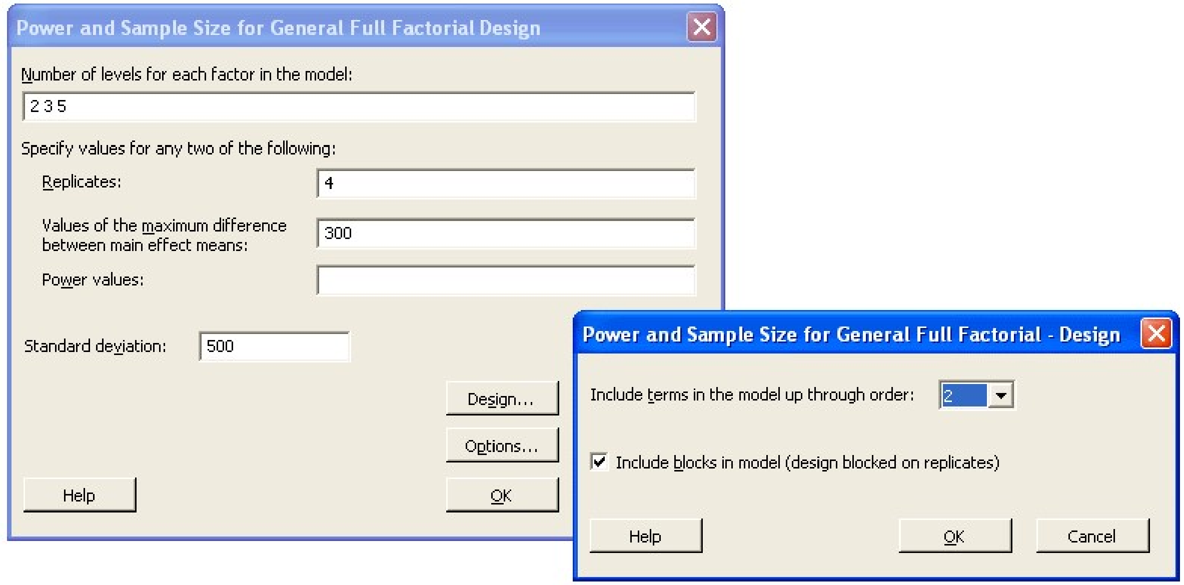
Balanced Full Factorial Design with Fixed Effects

Solution: Using MINITAB **Stat**> **Power and Sample Size**> **General Full Factorial Design**:

```
MTB > Power;
SUBC>  FDesign;
SUBC>   NLevels 2 3 5;
SUBC>   Reps 4;
SUBC>   MaxDifference 300;
SUBC>   Sigma 500;
SUBC>   TOrder 2;
SUBC>   FitB;
SUBC>   Alpha 0.05;
SUBC>   GPCurve.
```

Power and Sample Size

```
General Full Factorial Design

Alpha = 0.05  Assumed standard deviation = 500

Factors: 3  Number of levels: 2, 3, 5

Include terms in the model up through order:  2
Include blocks in model.


   Maximum        Total
Difference  Reps   Runs     Power
       300     4    120  0.328605
```



Fixed Effects in Mixed Models

- For a fixed variable A, the hypotheses to be tested are $H_0 : \mu_i = \mu_j$ for all pairs of A levels versus $H_A : \mu_i \neq \mu_j$ for at least one pair of A levels.
- The test is performed with the F statistic

$$F_A = \frac{MS_A}{MS_{\epsilon(A)}}$$

where $MS_{\epsilon(A)}$ is the mean square associated with the error for estimating the A effect.
- The power is given by

$$F_{1-\alpha} = F_{\beta,\phi_A}$$

where

$$\phi_A = \frac{N}{a} \frac{\sum_{i=1}^{a} \alpha_i^2}{MS_{\epsilon(A)}}$$

and N is the total number of observations.

Random Effects in Mixed Models

- The hypotheses to be tested are $H_0 : \sigma_R^2 = 0$ versus $H_A : \sigma_R^2 > 0$.
- The F statistic is

$$F_R = \frac{MS_R}{MS_{\epsilon(R)}}$$

where MS_R is the ANOVA mean square associated with R and $MS_{\epsilon(R)}$ is the mean square associated with the error term for testing the R effect.
- Under $H_A : \sigma_R^2 > 0$, the distribution of

$$F'_R = \frac{F_{1-\alpha}}{F_R}$$

follows the central F distribution with df_R numerator and $df_{\epsilon(R)}$ denominator degrees of freedom.

Random Effects in Mixed Models

- For specified values of the variances required to estimate MS_R and $MS_{\epsilon(R)}$ under H_A, the expected value of F_R is

$$E(F_R) = E\left(\frac{MS_R}{MS_{\epsilon(R)}}\right) \simeq \frac{E(MS_R)}{E(MS_{\epsilon(R)})}$$

and the corresponding power to reject H_0 is approximately

$$\pi \simeq P\left(\frac{F_{1-\alpha}}{E(F_R)} < F < \infty\right).$$

Fixed and Random Effects in Mixed Models

Example: A balanced full factorial experiment is to be performed using $a = 3$ levels of a fixed variable A, $b = 5$ randomly selected levels of a random variable B, and $n = 4$ replicates.

a) Determine the power to reject $H_0 : \alpha_i = 0$ *for all* i when the A-level biases are $\alpha_i = \{-20, 20, 0\}$ with $\sigma_B = 25$, $\sigma_{AB} = 0$, and $\sigma_\epsilon = 40$. Assume that the AB interaction term will be included in the ANOVA even though its expected variance component is 0.

b) Determine the power to reject $H_0 : \sigma_B^2 = 0$ when $\sigma_B = 25$, $\sigma_{AB} = 0$, and $\sigma_\epsilon = 40$. Retain the AB interaction term in the model even though its variance component is 0.

Fixed and Random Effects in Mixed Models

Solution: a) The error mean square used for testing the A effect (that is, the denominator of F_A) is

$$MS_{\epsilon(A)} = MS_{AB}$$
$$= \widehat{\sigma}_{\epsilon}^{2} + n\widehat{\sigma}_{AB}^{2}.$$

The noncentrality parameter is

$$\phi_A = \frac{N}{a}\frac{\sum_{i=1}^{a}\alpha_i^2}{MS_{\epsilon(A)}}$$
$$= \frac{3 \times 5 \times 4}{3}\frac{(-20)^2 + (20)^2 + (0)^2}{(40)^2 + 4(0)^2}$$
$$= 10.$$

With $df_A = 2$ and $df_{AB} = 8$,

$$F_{0.95} = 4.459 = F_{1-\pi,10.0}$$

which is satisfied by $\pi = 0.640$.

Fixed and Random Effects in Mixed Models

Solution: b) The expected F_B value is approximately

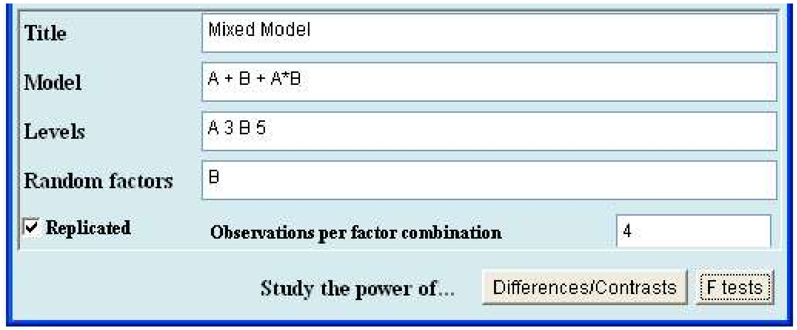
$$
\begin{aligned}
E(F_B) &\simeq \frac{E(MS_B)}{E(MS_{AB})} \\
&\simeq \frac{\sigma_\epsilon^2 + n\sigma_{AB}^2 + an\sigma_B^2}{\sigma_\epsilon^2 + n\sigma_{AB}^2} \\
&\simeq \frac{(40)^2 + 4(0)^2 + 3 \times 4 \times (25)^2}{(40)^2 + 4(0)^2} \\
&\simeq 5.69.
\end{aligned}
$$

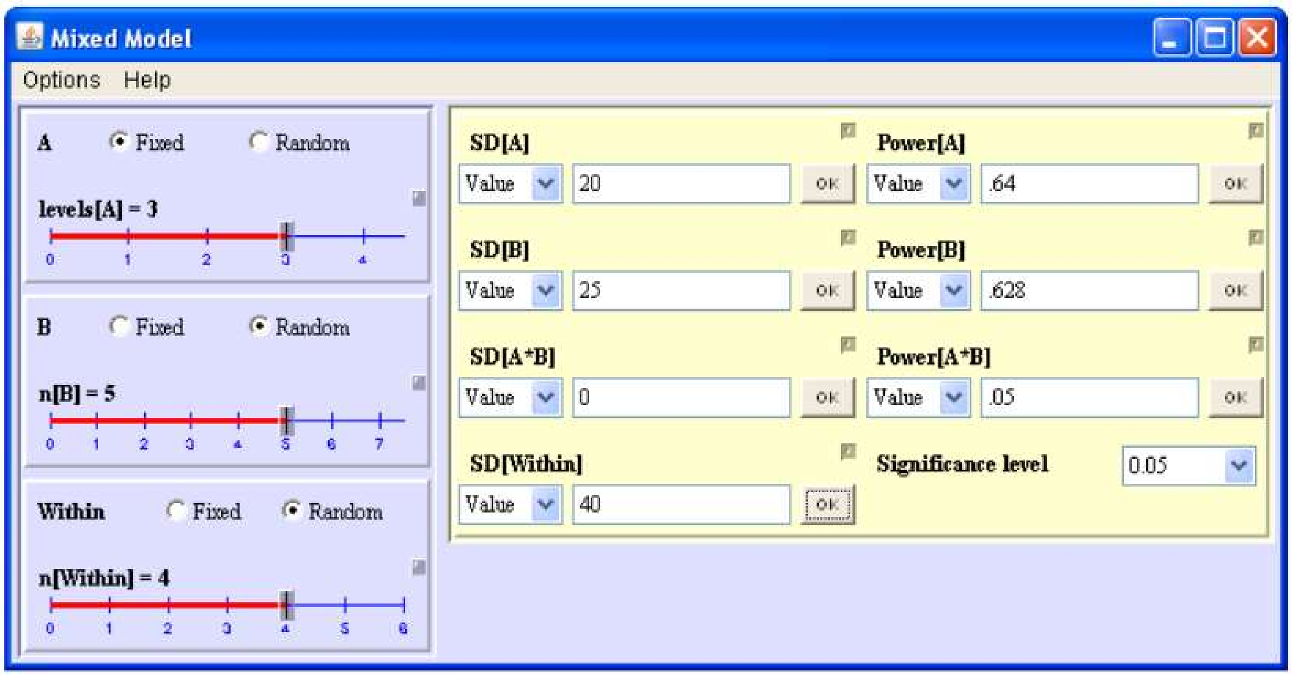
With $df_B = 4$ and $df_{AB} = 8$, the critical F value for the test for the B effect is $F_{0.95,4,8} = 3.838$, so the power is approximately

$$
\begin{aligned}
\pi &\simeq P\left(\frac{3.838}{5.69} < F < \infty\right) \\
&\simeq P(0.675 < F < \infty) \\
&\simeq 0.618.
\end{aligned}
$$

Fixed and Random Effects in Mixed Models

Solution: With $s_A = \sqrt{\left((-20)^2 + (20)^2 + (0)^2\right)/(3-1)} = 20$:





Two-Level Factorial Designs

- There are two goals in two-level factorial designs:
  - Detecting signficant effects (ANOVA)
  - Quantifying a regression coefficient (regression)
- For the purpose of testing for significant effects, use the balanced full factorial power calculation method. The total number of observations required can be approximated from

$$n2^k \simeq 4(t_{\alpha/2} + t_\beta)^2 \left(\frac{\sigma_\epsilon}{\delta}\right)^2.$$

  (Notice that the right hand side is almost constant!)
- For the purpose of estimating the regression coefficient associated with a variable, use the linear regression slope method, which gives:

$$n \geq \frac{1}{2^k}\left(\frac{t_{\alpha/2}\widehat{\sigma}_\epsilon}{\delta}\right)^2.$$

Two-Level Factorial Designs

Example: A two-level factorial experiment is limited to 32 experimental runs. Determine the power to detect an effect $\delta = \sigma_\epsilon$ using designs with one to five variables and all of the available resources. Assume that the model will include only main effects and two-factor interactions.

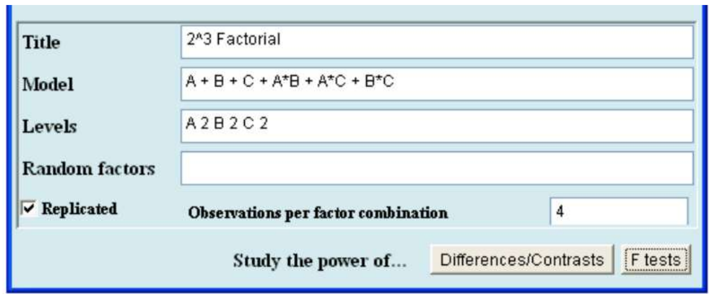
Solution: The table below shows the exact power as a function of the number of variables in the experiment.

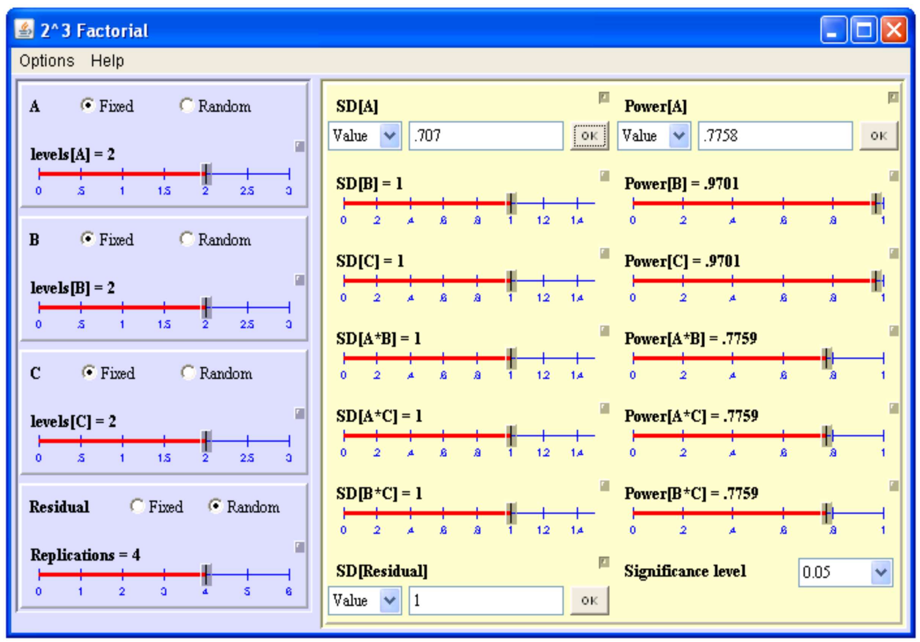
k	n	$n2^k$	π
1	16	32	0.781
2	8	32	0.779
3	4	32	0.776
4	2	32	0.767
5	1	32	0.757
6	$\frac{1}{2}$	32	0.757

Two-Level Factorial Designs

Solution: For the 2^3 design with

$$s_A = \sqrt{\left((-\tfrac{1}{2})^2 + (\tfrac{1}{2})^2\right)/(2-1)} = 0.707:$$





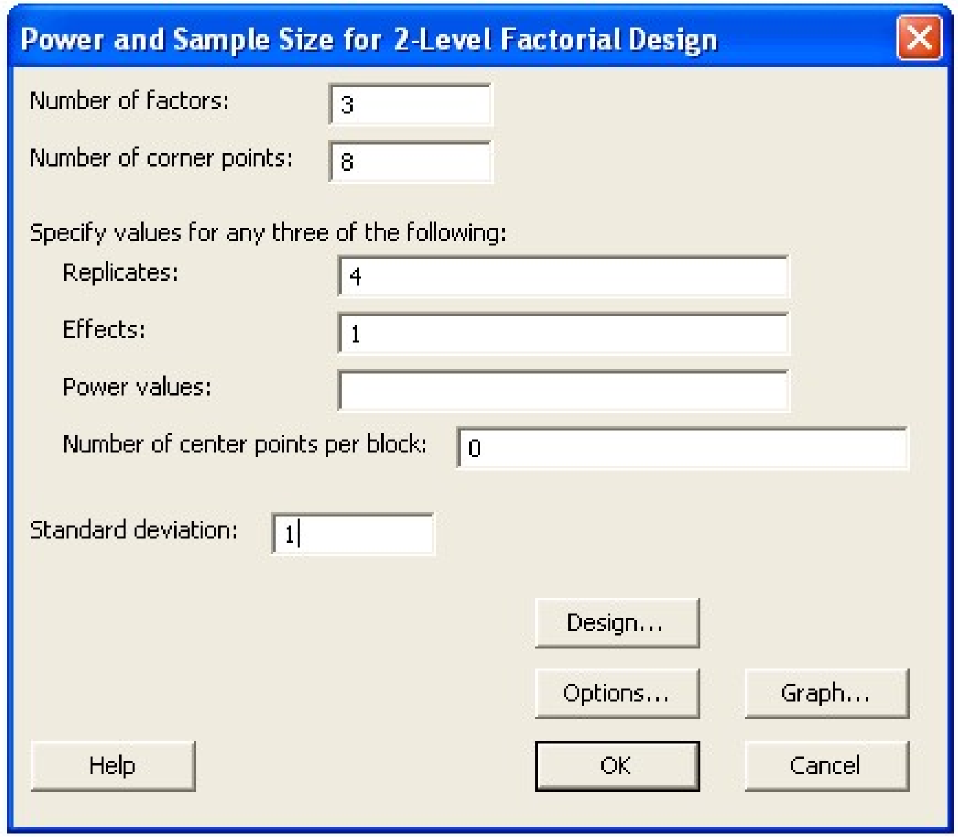
Two-Level Factorial Designs

Solution: Using MINITAB **Stat**> **Power and Sample Size**> **2-Level Factorial**:

```
MTB > Power;
SUBC>   FFDesign 3 8;
SUBC>     Reps 4;
SUBC>     Effect 1;
SUBC>     CPBlock 0;
SUBC>     Sigma 1;
SUBC>     Omit 1;
SUBC>     FitC;
SUBC>     FitB;
SUBC>   GPCurve.
```

Power and Sample Size

```
2-Level Factorial Design

Alpha = 0.05  Assumed standard deviation = 1

Factors:    3   Base Design: 3, 8
Blocks:   none

Number of terms omitted from model: 1


Center                 Total
Points  Effect  Reps    Runs     Power
     0       1     4      32  0.775898
```



Two-Level Factorial Designs

Example: How many replicates of a 2^3 design are required to determine the regression coefficient for a main effect with precision $\delta = 300$ with 95% confidence when the standard error of the model is expected to be $\sigma_\epsilon = 600$?

Solution: If the error degrees of freedom are sufficiently large that $t_{0.025} \simeq z_{0.025}$ then

$$\begin{aligned} n &\geq \frac{1}{2^3}\left(\frac{1.96 \times 600}{300}\right)^2 \\ &\geq 2. \end{aligned}$$

With only $2 \times 2^3 = 16$ total runs, the $t_{0.025} \simeq z_{0.025}$ assumption is not satisfied. Another iteration shows that the transcendental sample size condition is satisfied for $n = 3$ replicates of the 2^3 design.

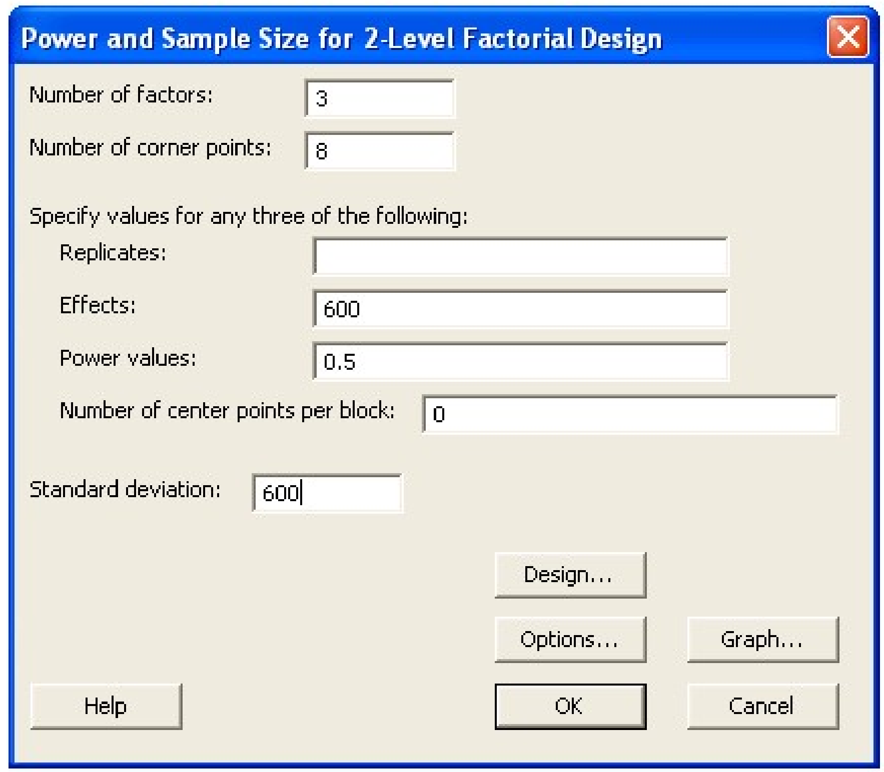
Two-Level Factorial Designs

Solution: Using MINITAB **Stat**> **Power and Sample Size**> **2-level Factorial Design** (Note: MINITAB's menu is expressed in terms of the effect size which is two times the value of the regression coefficient.):

```
MTB > Power;
SUBC>   FFDesign 3 8;
SUBC>     Effect 600;
SUBC>     Power 0.5;
SUBC>     CPBlock 0;
SUBC>     Sigma 600;
SUBC>     Omit 1;
SUBC>     FitC;
SUBC>     FitB;
SUBC>   GPCurve.
```

Power and Sample Size

```
2-Level Factorial Design

Alpha = 0.05  Assumed standard deviation = 600

Factors:    3   Base Design: 3, 8
Blocks:   none

Number of terms omitted from model: 1


Center                Total  Target
Points  Effect  Reps   Runs   Power  Actual Power
     0     600     3     24     0.5      0.636714
```



Binary Responses in 2^k Designs

- The observation that the total number of runs is almost invariant with respect to k in 2^k designs with quantitative responses extends to binary responses.
- For a 2^k design with a binary response, calculate the total sample size using the two proportions method and then distribute the observations uniformly over all of the cells of the 2^k design.

Binary Responses in 2^k Designs

Example: Determine the number of replicates required for a 2^3 design with a binary reponse if the experiment should reject $H_0 : p = 0.02$ with 90% power when $p = 0.05$.

Solution: Using MINITAB **Stat**> **Power and Sample Size**> **2 Proportions**, the number of observations required is $n = 787$ per treatment group or $2 \times 787 = 1574$ in total. There are $2^3 = 8$ cells in the experiment, so the number of observations per cell should be $1574/8 \simeq 197$.

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. **Reliability**
    a. **Reliability Parameter Estimation**
    b. **Reliability Demonstration Tests**
    c. **Two-sample Reliability Tests**
    d. **Interference**
10. Statistical Quality Control
11. Resampling Methods

How are Reliability/Survival Statistical Methods Different from Classical Statistical Methods?

- Responses are often, but aren't limited to, time or number of cycles to failure
- Some additional distributions: exponential, Weibull, ...
- Censored observations:
  - Right censored - the experiment is suspended before a unit fails
  - Left censored - a unit fails before the first time it is observed
  - Interval censored - a unit fails between two observation times
- Analysis methods: Special methods are required for
  - Non-normal error distributions
  - Censored observations

How are Reliability/Survival Statistical Methods Similar to Classical Statistical Methods?

- Reliability/Survival methods also involve point estimates, confidence intervals, and hypothesis tests
- Reliability/Survival methods also involve issues of distribution location, variation, and shape
- Reliability/Survival experiments are also available for one sample, paired samples, two samples, and many samples

Sample Sizes for Reliability

- **The estimation precision for reliability parameters is determined by the number of failures, not by the number of units tested.**
- In the methods that follow:
  - The symbol n indicates the number of units tested
  - The symbol r indicates the number failures
  - Easy to use normal approximations will be used instead of the more accurate but more complicated methods
- Sample size calculations for reliability problems can be performed for:
  - Parameters, e.g. , exponential mean μ, Weibull shape β or scale η, ...
  - Percentiles, e.g. the time at which a specified fraction (percent) of the population fails, e.g. $B1$, $B10$, $LD50$, ...
  - Percent/failure fraction/reliability at a specified time

Confidence Interval for the Exponential MTTF

Under the exponential reliability model the reliability at time t is

$$R(t;\mu) = e^{-t/\mu}$$

where μ is the mean time to failure (MTTF). A point estimate for μ is

$$\bar{t} = \frac{1}{r}\sum_{i=1}^{n} t_i.$$

When r is large an approximate confidence interval for μ is given by

$$P(\bar{t}(1-\delta) < \mu < \bar{t}(1+\delta)) = 1-\alpha$$

where the confidence interval relative half-width is

$$\delta = \frac{z_{\alpha/2}}{\sqrt{r}}.$$

Then the approximate number of failures required to obtain a specified CI half-width is

$$r = \left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$

Confidence Interval for the Exponential MTTF

Example: How many units must be tested to failure to determine the exponential mean life μ with 20% precision and 95% confidence?

Solution: The goal of the experiment is to determine a confidence interval for μ of the form

$$P(\bar{t}(1-\delta) < \mu < \bar{t}(1+\delta)) = 1-\alpha$$

With $\alpha = 0.05$ and $\delta = 0.2$, the required number of failures is

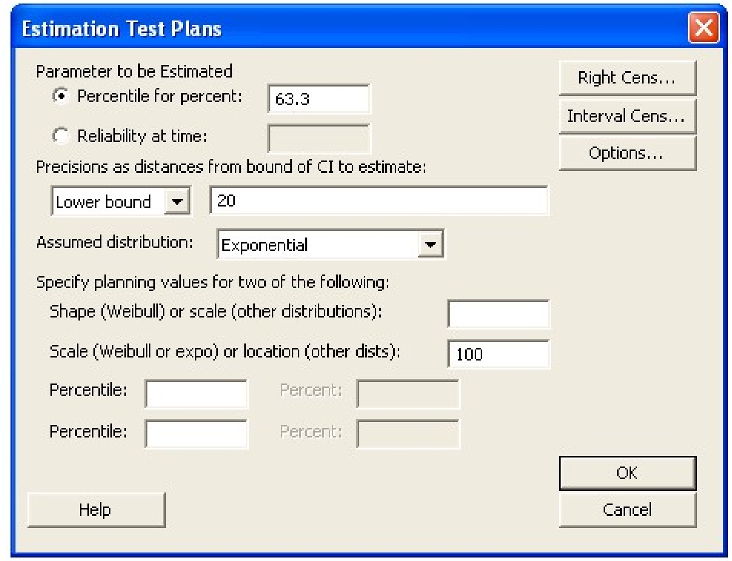
$$\begin{aligned} r &= \left(\frac{1.96}{0.20}\right)^2 \\ &= 97. \end{aligned}$$

Confidence Interval for the Exponential MTTF

Solution: Using MINITAB **Stat**> **Reliability/Survival**> **Test Plans**> **Estimation**:

- Specify the exponential distribution
- MINITAB is referenced by failure, not survival, rates, so …
- The desired percentile is μ with corresponding percent $1 - e^{-t/\mu}\big|_{t=\mu} = 1 - e^{-1} = 0.632.$
- Specify the confidence interval half-width.
- Average the sample sizes for the lower and upper bounds solutions.

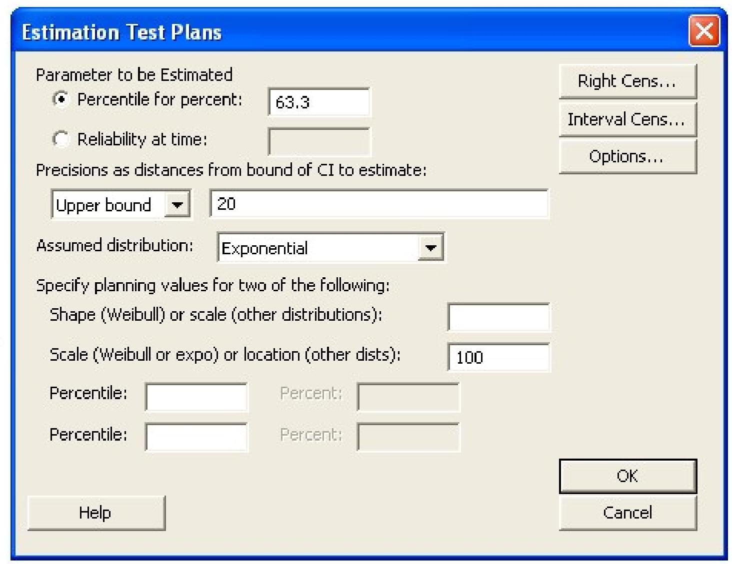
```
MTB > Etestplan;
SUBC>   EPtile 63.3;
SUBC>   Dlower 20;
SUBC>   Exponential;
SUBC>   ScLocation 100;
SUBC>   TwoSided.
```



Estimation Test Plans

```
Uncensored data

Estimated parameter: 63.3th percentile
Calculated planning estimate = 100.239
Target Confidence Level = 95%
Precision in terms of the lower bound of a two-sided confidence interval.

Planning distribution: Exponential
Scale = 100


                        Actual
             Sample  Confidence
Precision      Size       Level
       20        78     95.0641
```

```
MTB > Etestplan;
SUBC>   EPtile 63.3;
SUBC>   Dupper 20;
SUBC>   Exponential;
SUBC>   ScLocation 100;
SUBC>   TwoSided.
```



Estimation Test Plans

```
Uncensored data

Estimated parameter: 63.3th percentile
Calculated planning estimate = 100.239
Target Confidence Level = 95%
Precision in terms of the upper bound of a two-sided confidence interval.

Planning distribution: Exponential
Scale = 100


                        Actual
             Sample  Confidence
Precision      Size       Level
       20       117     95.0909
```

Confidence Interval for an Exponential Percentile

The sample size calculation for the confidence interval for the exponential mean also applies to all other percentiles.

Example: How many units must be tested to failure to determine, with 20% precision and 95% confidence, any failure percentile under the assumption that the reliability distribution is exponential?

Solution: The goal of the experiment is to determine a 95% confidence interval for the $100f^{th}$ failure percentile t_f of the form

$$P\left(\widehat{t}_f(1-\delta) < t_f < \widehat{t}_f(1+\delta)\right) = 0.95$$

where

$$\hat{t}_f = -\bar{t}\ln(1-f).$$

With $\alpha = 0.05$ and $\delta = 0.2$, the required number of failures is

$$r = \left(\frac{1.96}{0.20}\right)^2 = 97.$$

Confidence Interval for an Exponential Reliability

An approximate large-sample $(1-\alpha)100\%$ confidence interval for the exponential reliability $R(t;\mu)$ is

$$P\left(\hat{R}(1-\delta) < R < \hat{R}(1+\delta)\right) = 1-\alpha$$

where the confidence interval's relative half-width is

$$\delta = -\frac{z_{\alpha/2}\ln\left(\hat{R}\right)}{\sqrt{r}}.$$

Then the number of failures r required to obtain a specified confidence interval half-width is

$$r = \left(\frac{z_{\alpha/2}\ln\left(\hat{R}\right)}{\delta}\right)^2.$$

Confidence Interval for an Exponential Reliability

Example: How many units must be tested to failure in an experiment to determine, with 95% confidence, the exponential reliability to within 10% of its true value if the expected reliability is 80%?

Solution: With $\alpha = 0.05$, $\delta = 0.10$, and $\widehat{R} = 0.80$, the required number of failures is

$$r = \left(\frac{z_{\alpha/2} \ln\left(\widehat{R}\right)}{\delta} \right)^2$$

$$= \left(\frac{1.96 \ln(0.80)}{0.10} \right)^2$$

$$= 20.$$

Confidence Interval for the Weibull Scale Parameter

The Weibull reliability at time t is given by

$$R(t; \eta, \beta) = e^{-(t/\eta)^{\beta}}$$

where η is the scale factor and β is the shape factor. When β is known but η is not, which is often the case, then after the variable transformations $t' = t^{\beta}$ and $\eta' = \eta^{\beta}$ the Weibull distribution is transformed into the exponential distribution and the results from that method apply. This leads to the the approximate confidence interval for the scale parameter

$$P(\widehat{\eta}(1-\delta) < \eta < \widehat{\eta}(1+\delta)) = 1-\alpha$$

where the confidence interval's relative half-width is

$$\delta = \frac{z_{\alpha/2}}{\beta\sqrt{r}}.$$

Then the number of failures required to obtain specified relative precision δ is

$$r = \left(\frac{z_{\alpha/2}}{\beta\delta}\right)^2.$$

Confidence Interval for the Weibull Scale Parameter

Example: How many units must be tested to failure to estimate, with 20% precision and 95% confidence, the Weibull scale factor if the shape factor is known to be $\beta = 2$?

Solution: The goal of the experiment is to obtain a confidence interval for the Weibull scale factor with $\delta = 0.20$ and $\alpha = 0.05$. The required number of failures is

$$\begin{aligned} r &= \left(\frac{z_{\alpha/2}}{\beta\delta}\right)^2 \\ &= \left(\frac{1.96}{2 \times 0.20}\right)^2 \\ &= 25. \end{aligned}$$

Confidence Interval for the Weibull Shape Parameter

A $(1-\alpha)100\%$ confidence interval for the Weibull shape parameter β is required of the form

$$P\left(\hat{\beta}(1-\delta) < \beta < \hat{\beta}(1+\delta)\right) = 1-\alpha$$

where δ is the relative precision of the β estimate given by

$$\delta = \frac{z_{\alpha/2}}{\pi}\sqrt{\frac{6}{r}}.$$

Then the number of failures required to obtain a specified relative precision for the β estimate is

$$r = 6\left(\frac{z_{\alpha/2}}{\pi\delta}\right)^2.$$

Confidence Interval for the Weibull Shape Parameter

Example: How many units must be tested to failure to estimate, with 95% confidence, the Weibull shape parameter to within 20% of its true value?

Solution: The goal of the experiment is to produce a 95% confidence interval for β with relative half-width $\delta = 0.20$. With $\alpha = 0.05$, the required number of failures is

$$
\begin{aligned}
r &= 6\left(\frac{z_{\alpha/2}}{\pi\delta}\right)^2 \\
&= 6\left(\frac{1.96}{\pi \times 0.20}\right)^2 \\
&= 59.
\end{aligned}
$$

Confidence Interval for a Weibull Percentile

Just as the confidence interval for the exponential reliability mean μ is a special case of the confidence interval for the failure percentile, the confidence interval for the Weibull scale parameter η is a special case of the confidence interval for the Weibull failure percentile.

When the Weibull shape parameter β is known but the scale parameter η is unknown, which is often the case, an approximate confidence interval for the $100f^{th}$ failure percentile t_f is

$$P(\hat{t}_f(1-\delta) < t_f < \hat{t}_f(1+\delta)) = 1-\alpha$$

where

$$\delta = \frac{z_{\alpha/2}}{\beta\sqrt{r}}.$$

This is the same confidence interval half-width as was obtained for the Weibull scale parameter.

Confidence Interval for a Weibull Reliability

When the number of failures r is large, an approximate confidence interval for the Weibull reliability has the form

$$P\left(\widehat{R}(1-\alpha) < R < \widehat{R}(1+\delta)\right) = 1-\alpha$$

where the confidence interval half-width is

$$\delta = \frac{z_{\alpha/2}}{\sqrt{r}}\left(\frac{1-\widehat{R}}{\widehat{R}}\right).$$

Then the number of failures required to determine the reliability with relative precision δ is

$$r = \left(\frac{z_{\alpha/2}}{\delta}\left(\frac{1-\widehat{R}}{\widehat{R}}\right)\right)^2.$$

Confidence Interval for a Weibull Reliability

Example: How many units must be tested to failure to determine the Weibull reliability with 5% precision and 95% confidence at a time when the reliability is expected to be 90%?

Solution: The required number of failures is

$$r = \left(\frac{z_{\alpha/2}}{\delta} \left(\frac{1 - \hat{R}}{\hat{R}} \right) \right)^2$$

$$= \left(\frac{1.96}{0.05} \left(\frac{1 - 0.9}{0.9} \right) \right)^2$$

$$= 19.$$

Reliability Demonstration Tests

A *reliability demonstration test* is performed by putting n units on test for time t and observing the number of failures that occur within that time, i.e. the test is time-terminated. In order to demonstrate that the exponential mean life μ exceeds a specified value μ_0 with confidence $1 - \alpha$, that is:

$$P(\mu_0 < \mu < \infty) = 1 - \alpha$$

the test parameters must meet the condition:

$$b(c = r; n, p) \leq \alpha$$

where $b(c; n, p)$ is the cumulative binomial distribution and the probability of failure at time t is

$$p = 1 - R(t; \mu_0) = 1 - e^{-t/\mu_0}$$

Demonstration Test

Example: Determine the number of units that must be put on test for 200 hours without any failures to show that the MTTF of a system exceeds 400 hours with 95% confidence. Assume that the life distribution is exponential and that the test is time terminated.

Solution: The goal of the experiment is to determine the value of n with $r = 0$ failures in $t = 200$ hours of testing such that:

$$P(400 < \mu < \infty) = 0.95$$

With $\mu_0 = 400$ the $t = 200$ hour reliability is:

$$R(t = 200; \mu_0 = 400) = e^{-\frac{200}{400}} = 0.6065$$

so the probability that a unit will fail before 200 hours is $p = 1 - 0.6065 = 0.3935$. With $r = 0$ and $\alpha = 0.05$ the smallest value of n that meets the condition:

$$b(0; n, 0.3935) \leq 0.05$$

is $n = 6$ since:

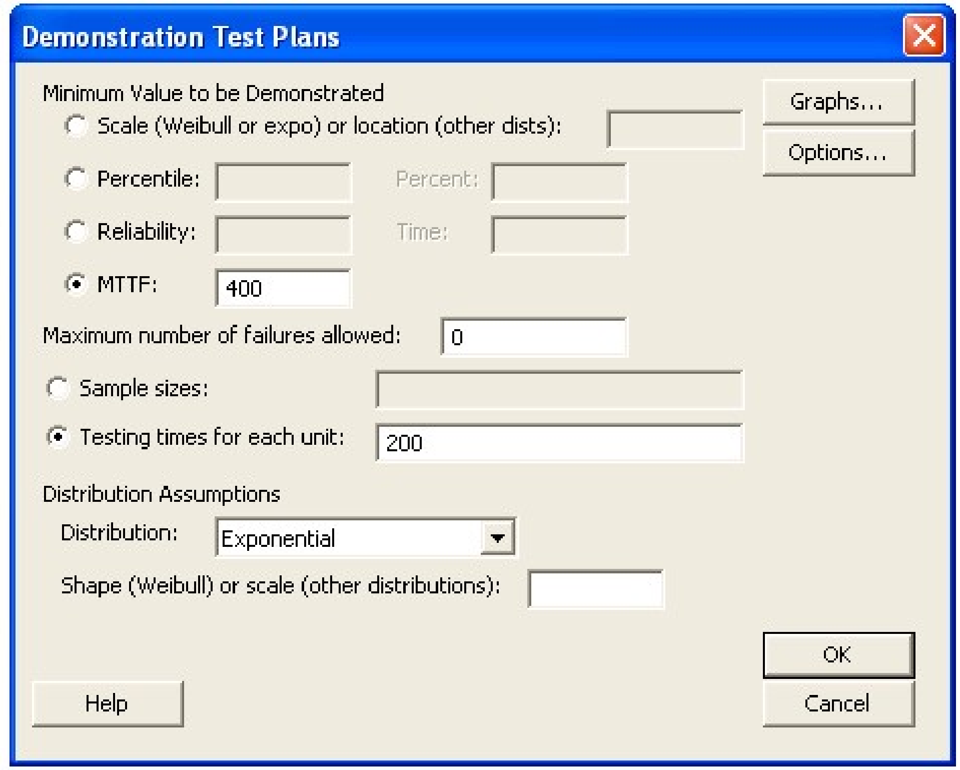
$$b(0; 6, 0.3935) = 0.04977$$

Using MINITAB **Stat**> **Reliability/Survival**> **Test Plans**> **Demonstration**:

```
MTB > DtestPlan 0;
SUBC>   MTTF 400 ;
SUBC>   TTime 200 ;
SUBC>   Exponential;
SUBC>   GPOPGraph.
```

Demonstration Test Plans

```
Reliability Test Plan
Distribution: Exponential
MTTF Goal = 400, Target Confidence Level = 95%


                                      Actual
Failure  Testing  Sample  Confidence
   Test     Time    Size       Level
      0      200       6     95.0213
```



Demonstration Test

Example: Determine the number of units that must be put on test for 10000 hours without any failures to demonstrate 90% reliability at 12000 hours with 95% confidence. Assume that the life distribution is Weibull with $\beta = 2.2$.

Solution: Following Mathews, Sample Size Calculations, p. 206, Example 9.18: The goal of the experiment is to demonstrate 90% reliability at $t_0 = 12000$ hours with 95% confidence or

$$P(12000 < t_{0.10} < \infty) = 0.95$$

The units to be put on test will be operated for $t' = 10000$ hours and then the test will be suspended.

From Table 9.2 with $f_0 = 0.10$ and $\beta = 2.2$

$$\begin{aligned} f' &= 1 - (1 - f_0)^{(t'/t_0)^\beta} \\ &= 1 - (1 - 0.10)^{(10000/12000)^{2.2}} \\ &= 0.0681 \end{aligned}$$

The number of units that must be tested must satisfy the condition

$$b(r = 0; n, f') \leq \alpha$$

$$b(r = 0; n, 0.0681) \leq 0.05$$

which gives $n = 43$.

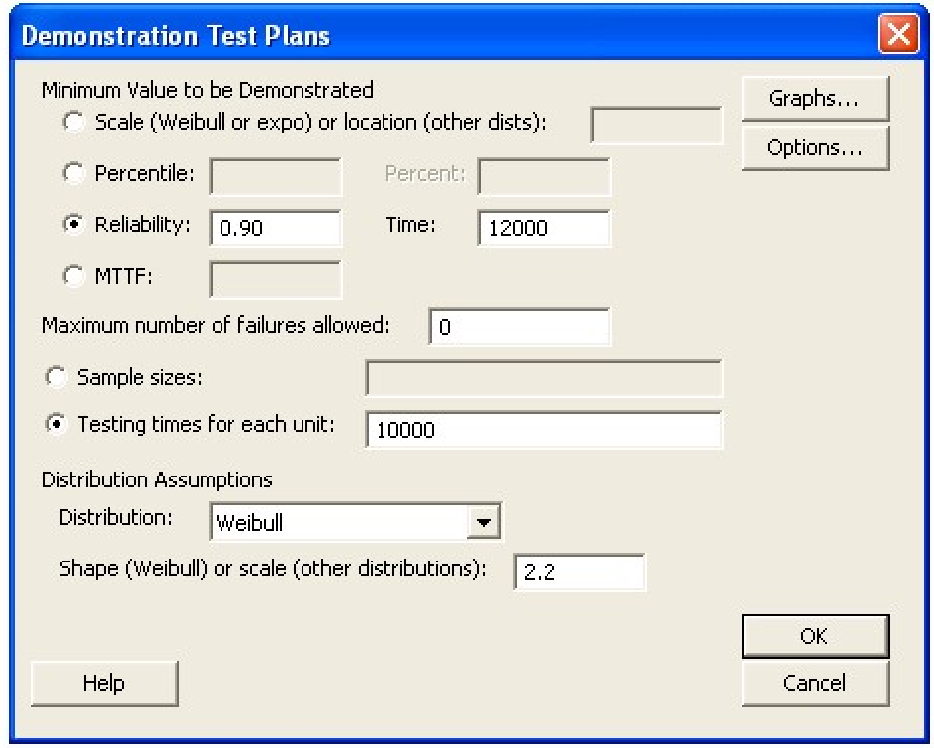
Using MINITAB **Stat**> **Reliability/Survival**> **Test Plans**> **Demonstration**:

```
MTB > DtestPlan 0;
SUBC>   Reliability 0.90;
SUBC>     Time 12000 ;
SUBC>   TTime 10000 ;
SUBC>   Weibull;
SUBC>   ShScale 2.2;
SUBC>   GPOPGraph.
```

Demonstration Test Plans

```
Reliability Test Plan
Distribution: Weibull, Shape = 2.2
Reliability Goal = 0.9, Target Confidence Level = 95%
```

Failure Test	Testing Time	Sample Size	Actual Confidence Level
0	10000	43	95.1854



Two-Sample Reliablity Tests

- Two-sample reliability tests are used to test for differences between two independent reliability distributions.
- Such tests may be performed for reliability parameters, percentiles, and survival rates at a specified endpoint.
- The log-rank test is a popular two-sample reliability test for a difference in the survival rates between two treatment groups.

Two-Sample Log-Rank Test

- The hypotheses to be tested are $H_0 : h_1(t) = h_2(t)$ versus $H_A : h_1(t) > h_2(t)$ where $h_1(t)$ and $h_2(t)$ are the time-dependent hazard rates
- The hazard ratio $h_2(t)/h_1(t)$ must be constant with respect to time, i.e. must meet the *proportional hazards* assumption.
- The log-rank test hypotheses are usually redefined in terms of the log-hazard ratio, r, which is estimated from survival probabilities $s_1(t)$ and $s_2(t)$ at any common time t under the proportional hazards assumption

$$r = \frac{\ln(s_2(t))}{\ln(s_1(t))}$$

- The log-hazard ratio is usually determined from the end-of-test ($t = t'$) survival probabilities.
- The log-rank test hypotheses may be rewritten as $H_0 : r = 1$ versus $H_A : r < 1$ where H_A is constructed to reject H_0 when the treatment group's survival rate is significantly greater than the control group's survival rate.

Two-Sample Log-Rank Test

Two popular methods for calculating power and sample size are presented for the log-rank test:

- Schoenfeld's method
- Lachin's method

The two methods give nearly identical results for the equal-sample-size case but diverge slightly when the sample sizes are not equal. Lachin's method is preferred in the unequal-sample-size case because it is more conservative.

Two-Sample Log-Rank Test

The sample sizes required to reject $H_0 : r = 1$ versus $H_A : r < 1$ when $r = r_A$ with power $\pi = 1 - \beta$ is given by:

- Schoenfeld's method:

$$n_1 = n_2 = \left(\frac{z_\alpha + z_\beta}{\ln(r_A)}\right)^2 \left(\frac{1}{1 - s_1(t')} + \frac{1}{1 - s_2(t')}\right)$$

- Lachin's method:

$$n_1 = n_2 = \frac{(z_\alpha + z_\beta)^2}{2 - s_1(t') - s_2(t')} \left(\frac{1 + r_A}{1 - r_A}\right)^2$$

Two-Sample Log-Rank Test

Example: Determine how many units must be included in a study to compare the survival rates of two treatments using the log-rank test if the control treatment is expected to have about 20% survivors at the end of the study and the study should have 90% power to reject H_0 if the experimental treatment has 40% survivors at the end of the study. Assume that the hazard rates are proportional and that the sample sizes will be equal.

Solution: From the expected end-of-study conditions under H_A the log-hazard ratio is estimated to be

$$r_A \simeq \frac{\ln(0.40)}{\ln(0.20)} = 0.5693.$$

Two-Sample Log-Rank Test

Solution(**continued**): The required sample size by Schoenfeld's method is

$$n = \left(\frac{z_{\alpha} + z_{\beta}}{\ln(r_A)}\right)^2 \left(\frac{1}{1 - s_1(t')} + \frac{1}{1 - s_2(t')}\right)$$

$$= \left(\frac{z_{0.05} + z_{0.10}}{\ln(0.5693)}\right)^2 \left(\frac{1}{1 - 0.2} + \frac{1}{1 - 0.4}\right)$$

$$= 79$$

and by Lachin's method is

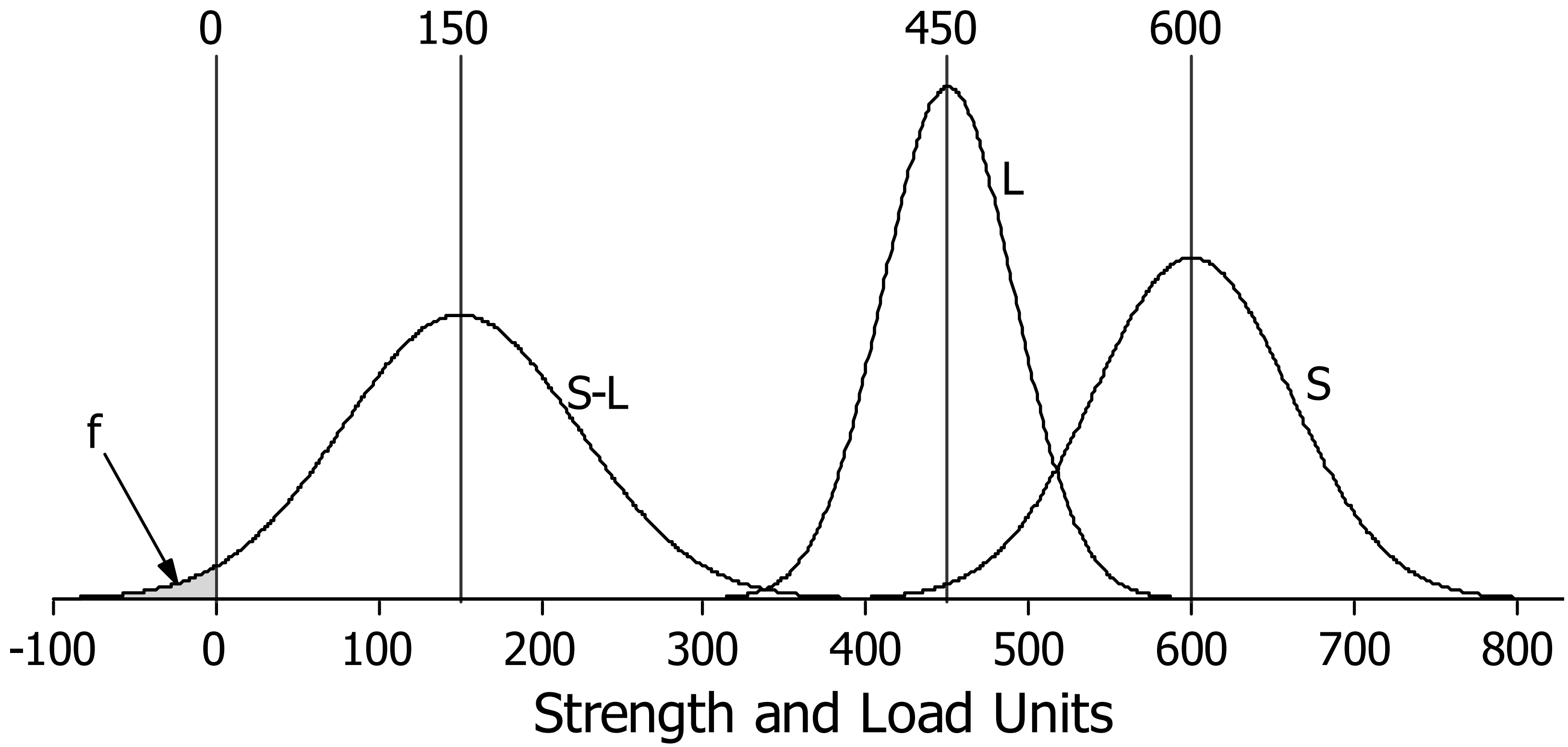
$$n = \frac{(z_{\alpha} + z_{\beta})^2}{2 - s_1(t') - s_2(t')} \left(\frac{1 + r_A}{1 - r_A}\right)^2$$

$$= \frac{(1.645 + 1.282)^2}{2 - 0.2 - 0.4} \left(\frac{1 + 0.5693}{1 - 0.5693}\right)^2$$

$$= 82$$

Interference

- Interference failures occur when a quality characteristic exceeds a limit where both the quality characteristic and the limit are statistically distributed.
- The figure shows a strength–load interference situation:



Distributions of strength, load, and their difference.

Interference

- If strength and load values are indicated with the symbol x, and if their probability density functions are given by $S(x)$ and $L(x)$, respectively, then the probability of interference failure is given by

$$f = \int_{-\infty}^{\infty} S(x_S)\left(\int_{x_S}^{\infty} L(x_L)dx_L\right)dx_S.$$

  The necessary integrations may be performed to solve for the failure probability when $S(x)$ and $L(x)$ are well defined; however, in many situations the actual interference analysis is performed by resampling from sample strength and load data.
- See Mathews for treatment of the normal-normal, exponential-exponential, and Weibull-Weibull interference failure sample size calculations.

Weibull-Weibull Interference

When the strength and load are both Weibull-distributed the approximate strength/load interference failure rate (f) is given by

$$f \simeq \left(\frac{\eta_L}{\eta_S}\right)^{\beta_S} \Gamma\left(1 + \frac{\beta_S}{\beta_L}\right)$$

where $\Gamma(\quad)$ is the gamma function. The approximation is satisfied when

$$\left(\frac{\eta_L}{\eta_S}\right)^{\beta_S} \ll 1,$$

which corresponds to small f - the usual condition of interest.

Weibull-Weibull Interference

Under the assumption that β_L and β_S are known, for large samples, the approximate one-sided upper confidence limit for f is given by

$$\Phi\left(0 < f < \widehat{f}_U\right) = 1 - \alpha$$

where $\widehat{f}_U = \widehat{f}(1+\delta)$ where

$$\delta = \frac{z_\alpha \beta_S}{\Gamma\left(1 + \frac{\beta_S}{\beta_L}\right)} \sqrt{\frac{1}{n_L \beta_L^2} + \frac{1}{n_S \beta_S^2}}\,.$$

For a specified value of the relative precision of the estimate δ in the equal-sample-size case $(n_L = n_S = n)$, this equation can be solved for the sample size to obtain

$$n = \left(\frac{z_\alpha}{\delta \Gamma\left(1 + \frac{\beta_S}{\beta_L}\right)}\right)^2 \left(1 + \frac{\beta_S^2}{\beta_L^2}\right).$$

Example: How many measurements of mating components in a device must be taken to demonstrate, with 95% confidence, that their true interference failure rate does not exceed the observed failure rate by 20% if the two distributions are known to be Weibull with $\beta_S = 2.5$ and $\beta_L = 1.5$?

Solution: The goal of the experiment is to acquire sufficient information to demonstrate the following one-sided upper confidence interval for the interference failure rate f:

$$P\left(0 < f < \widehat{f}(1 + 0.2)\right) = 0.95.$$

With $\delta = 0.2$ and $\alpha = 0.05$ we obtain the sample size

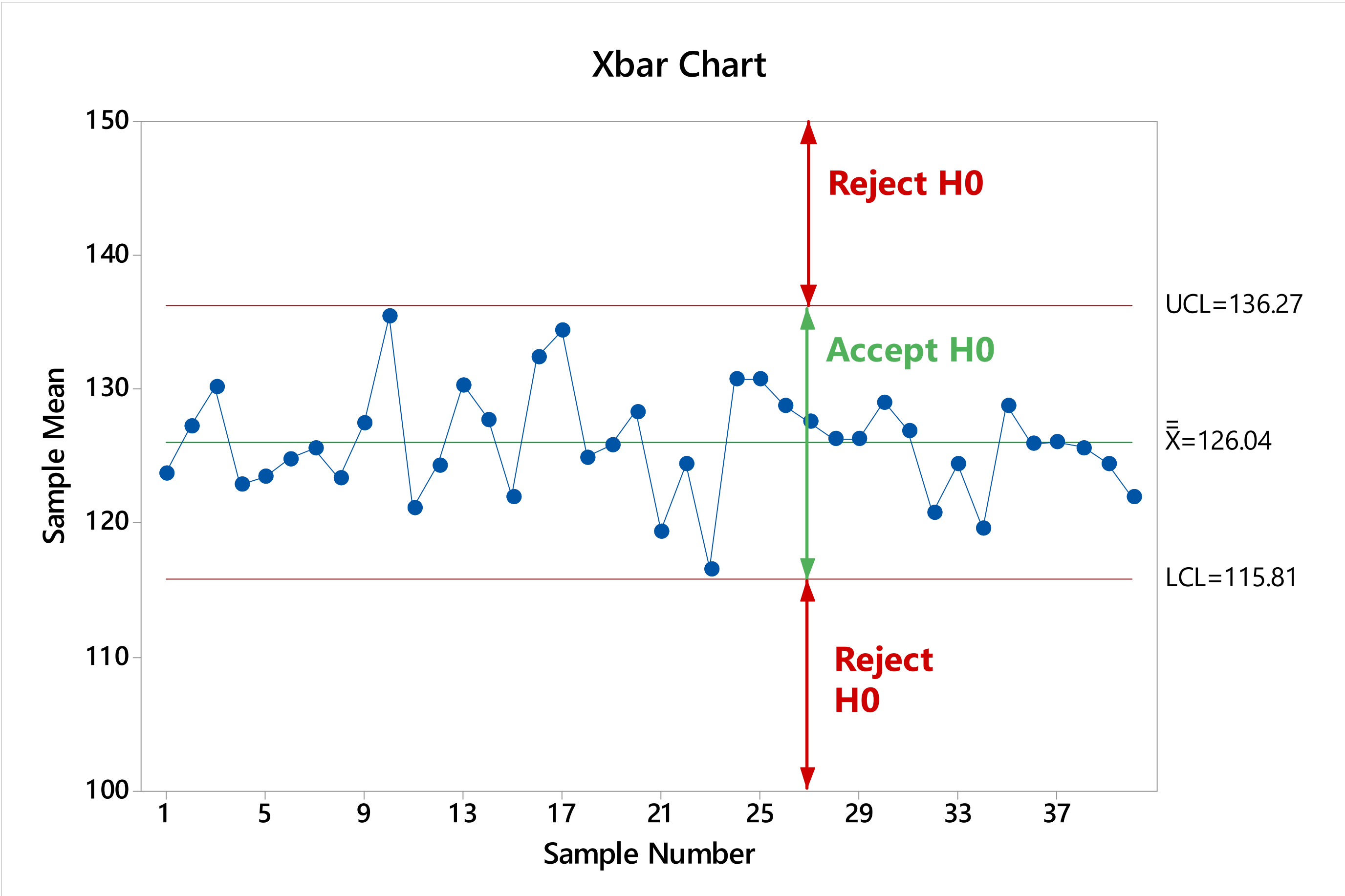
$$\begin{aligned} n &= \left(\frac{z_\alpha}{\delta \Gamma\left(1 + \frac{\beta_S}{\beta_L}\right)} \right)^2 \left(1 + \frac{\beta_S^2}{\beta_L^2}\right) \\ &= \left(\frac{1.645}{0.2 \times \Gamma(1 + \frac{2.5}{1.5})} \right)^2 \left(1 + \frac{2.5^2}{1.5^2}\right) \\ &= 113. \end{aligned}$$

Seminar Outline

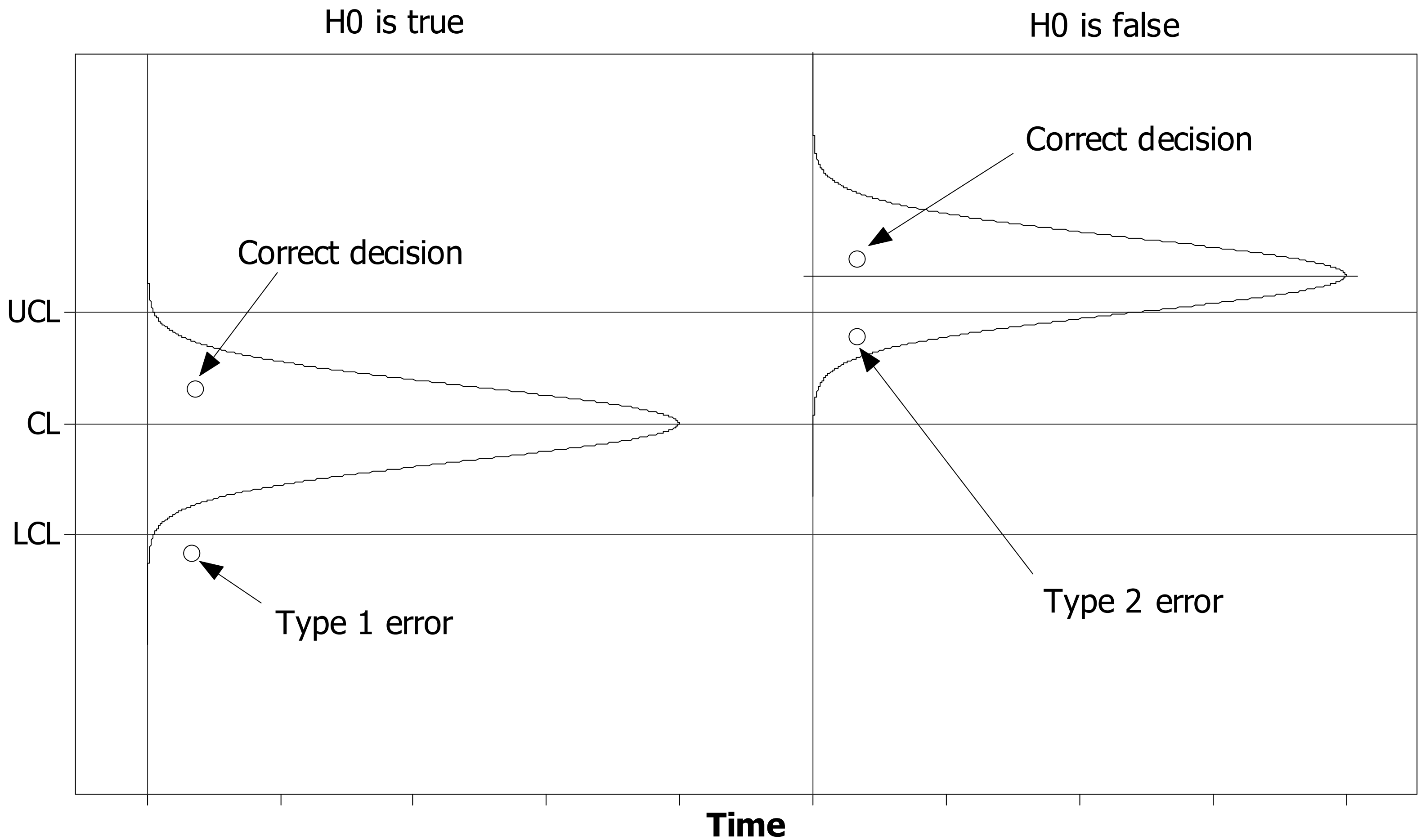
1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. **Statistical Quality Control**
    a. **SPC Charts**
    b. **Process Capability**
    c. **Tolerance Intervals**
    d. **Acceptance Sampling**
    e. **Gage Error Studies**
11. Resampling Methods

SPC Charts and Hypothesis Testing

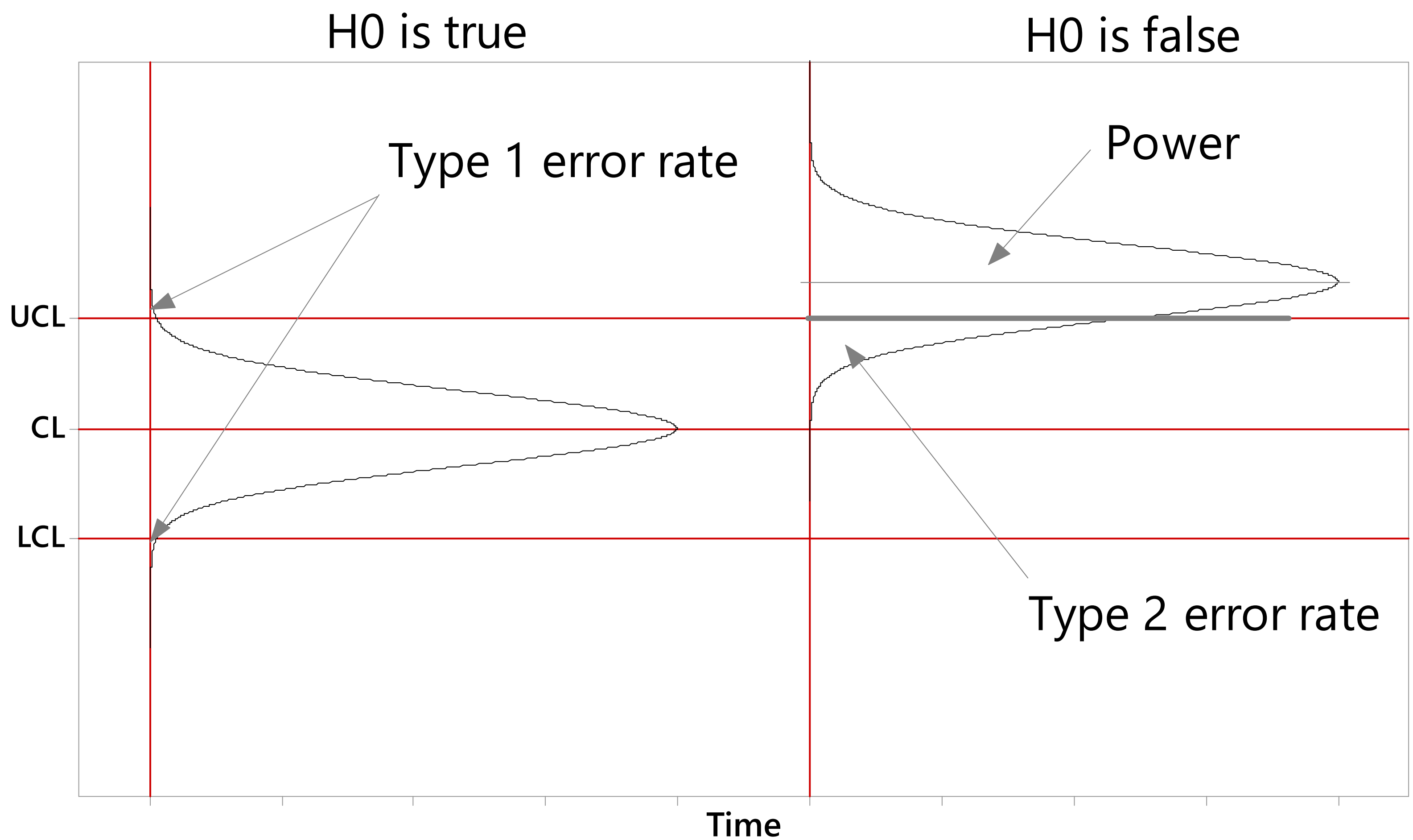
Shewhart's SPC charts provide a graphical hypothesis test of $H_0 : \mu = CL$ *(the process is in control)* versus $H_A : \mu \neq CL$ *(the process is out of control):*



Type 1 and Type 2 Errors in SPC Charts

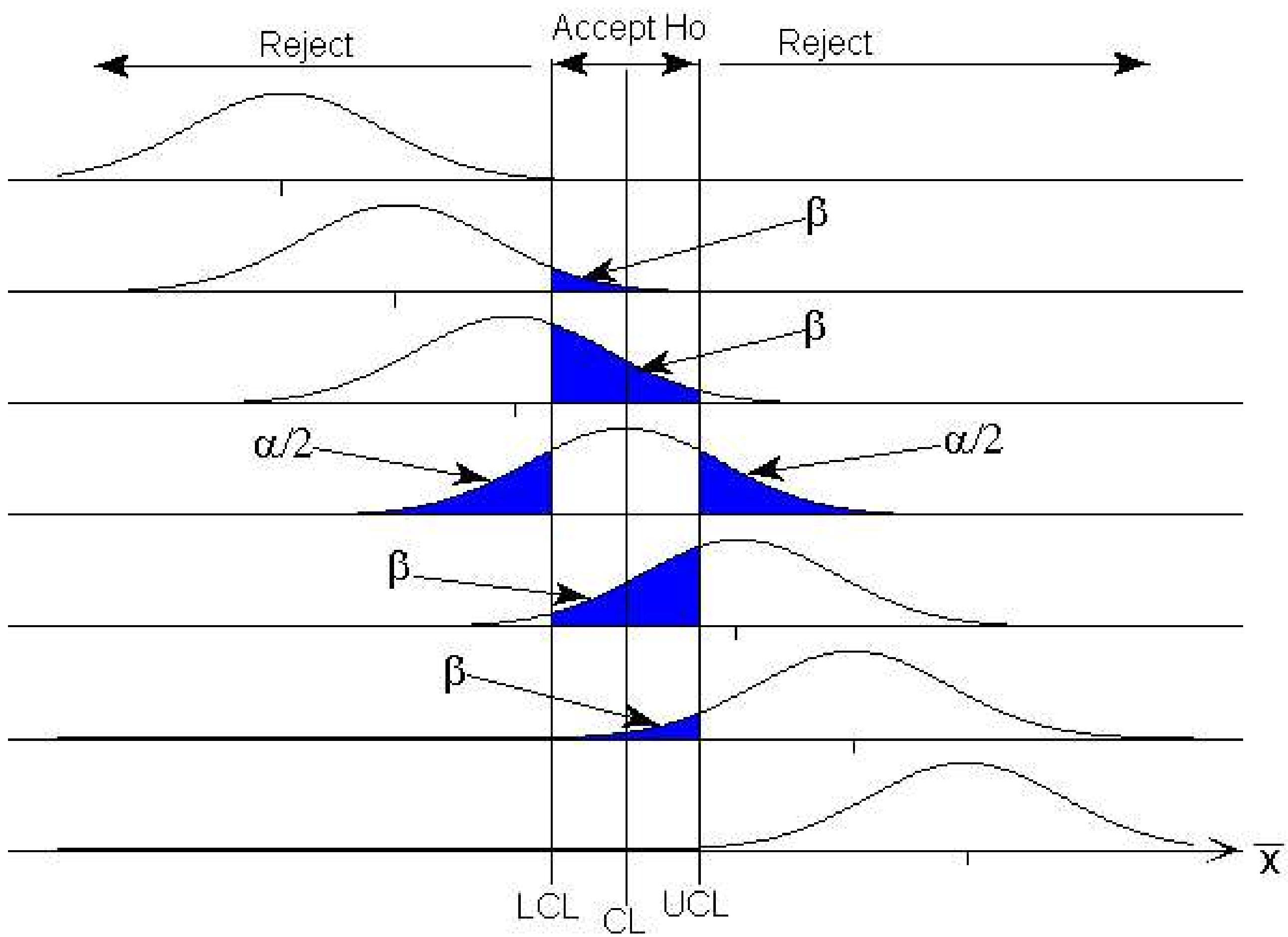


Type 1 and Type 2 Errors in SPC Charts



SPC Run Rule Power

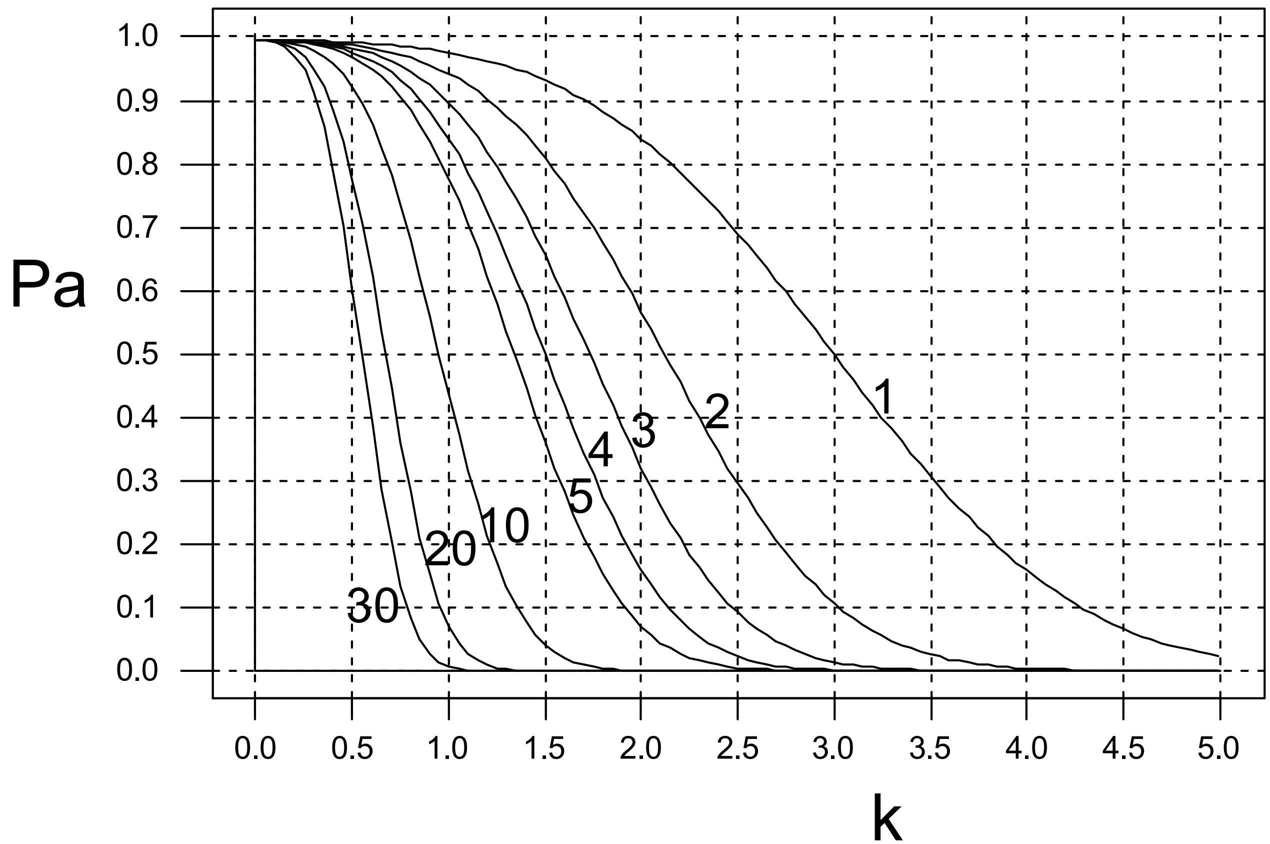
A run rule's power (its probability of detecting a shift in location) can be calculated as a function of the size of the shift.



SPC Run Rule Power

For the rule that the process is declared to be out of control if a single point falls beyond the usual 3σ contol limits applied to an $\bar{x}$ chart the rule's operating characteristic (OC) curve is given by

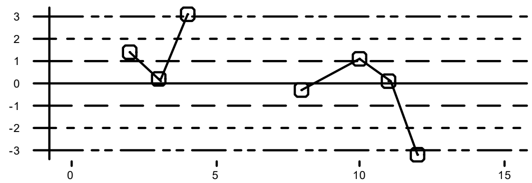
$$\Phi(LCL < \bar{x} < UCL; \mu \neq \mu_0, \sigma_{\bar{x}}) = \beta$$



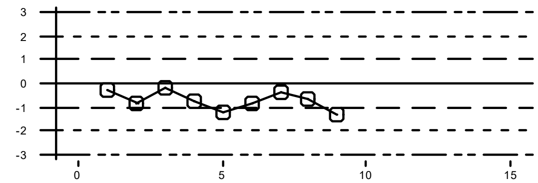
SPC Chart Run Rules

SPC charts are usually operated with several run rules:

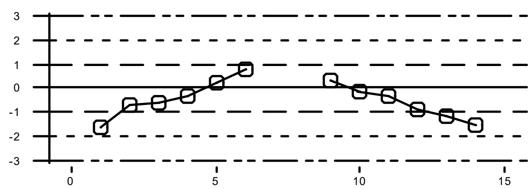
TEST 1: One point beyond 3σ



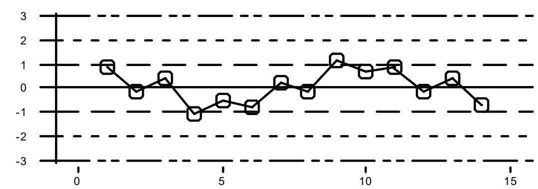
TEST 2: 9 points to one side of CL



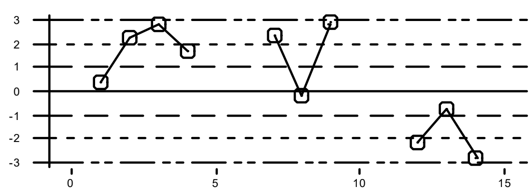
TEST 3: 6 point trend up or down



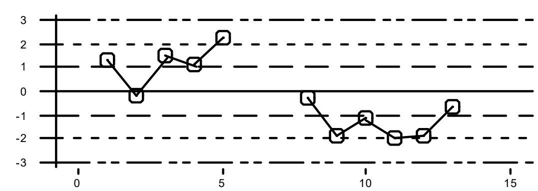
TEST 4: 14 points alternating up/down



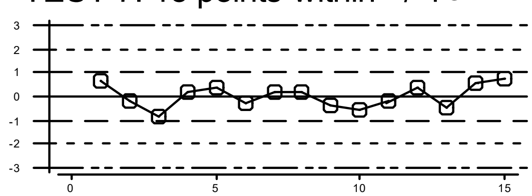
TEST 5: 2 of 3 points beyond 2σ



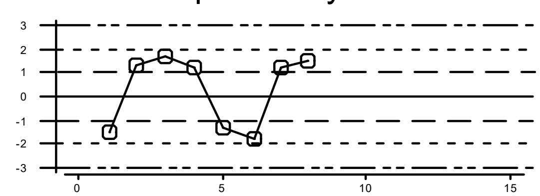
TEST 6: 4 out of 5 points beyond 1σ



TEST 7: 15 points within +/-1σ



TEST 8: 8 points beyond +/-1σ



Created by: Rebecca Malnar 9/12/99

Design of SPC Chart Run Rules

Good SPC chart run rules must meet the following conditions:

1. A rule must be easy to recognize on the chart
2. A rule must have a low false alarm / type 1 error rate
3. A rule must have a low missed alarm / type 2 error rate

Example: Evaluate the run rule: *If at least four of five consecutive points fall beyond* 1σ *to the same side of the center line then the process is out of control. (This is one of the Western Electric rules.)*

Solution: This pattern is easy to recognize on the chart so the first condition is met.

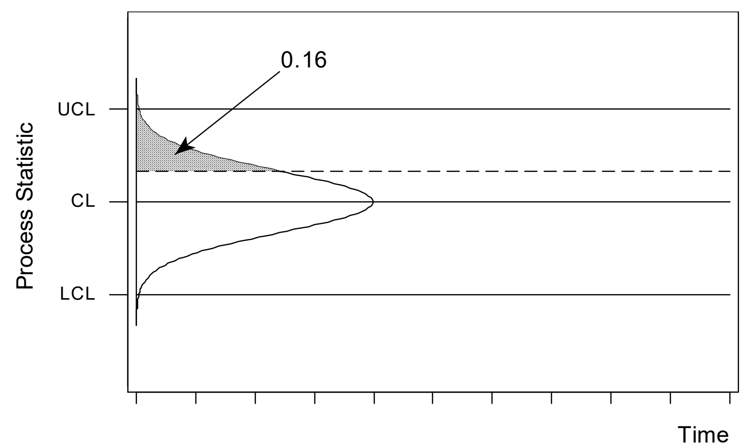
For the second condition, if the process is in control then the probability that any one point falls beyond 1σ of the centerline is:

$$\Phi(1 < z < \infty) = 0.16$$

Then the probability that at least four of five points fall beyond 1σ of the centerline is:

$$b(4 \le x \le 5; 5, 0.16) = 0.0029$$

Since this pattern can show up on either side of the chart we have $\alpha = 2(0.0029) = 0.0058$ which is acceptably low.



For the third condition, suppose the process shifts so the new process mean falls right on a control limit. Then the probability of a single point falling beyond 1σ of the center line is given by:

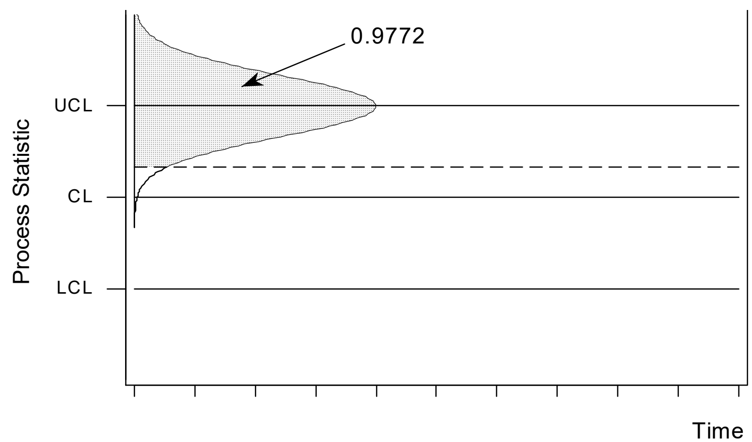
$$\Phi(-2 < z < \infty) = 0.9772$$

The probability that at least four of five consecutive points fall beyond 1σ of the center line is then:

$$b(4 \leq x \leq 5; 5, 0.9772) = 0.9950$$

This means that the probability of detecting the shift using this rule is about $1 - \beta = 0.9950$.

This rule meets all three conditions so it is a good run rule.



np and p Chart Sample Size

Rules for determining sample size for np and p charts:

1. To obtain $LCL \geq 0$:

$$n \geq \frac{9(1-p)}{p}$$

2. To detect a shift of the fraction defective from p to $p+\delta$ with 50% probability use:

$$n = \frac{9p(1-p)}{\delta^2}$$

3. To limit the frequency of zero defectives on the chart to less than 5% use:

$$n \geq \frac{3}{p}$$

Special Run Rule

The $n = 3/p$ sample size criterion is often paired with a special run rule: The process is out of control if two consecutive zeros occur on the chart. This is a good rule because it meets all of the requirements of a good run rule:

- It is easy to recognize
- Its type 1 error rate is reasonable ($\alpha = 0.05^2 = 0.0025$)
- The rule turns on hard when the process goes out of control low

c Chart Sample Size

1. To obtain a zero or positive lower control limit use a sampling unit that is large enough so that $c \geq 9$.
2. To detect a shift of the mean defect rate from c to $c + \delta$ with 50% probability use a sampling unit that is large enough so that $c \geq \delta^2/9$.
3. To limit the number of occurrences of zero defectives on the chart to less than 5% use:

$$c \geq 3$$

   This sample size criterion is often used with a special run rule: the process is out of control if two consecutive zeros occur on the chart.

Multiple Testing Errors in SPC

- SPC is a prime example of an opportunity to suffer from excessive type 1 / false alarm errors caused by multiple testing from keeping too many run rules on too many charts.
- Each run rule has its own false alarm / type 1 error rate.
- The run rules are not strictly independent of each other but their errors are roughly additive.
- The error rates from several charts are roughly additive.

Multiple Testing Errors in SPC

Example: Suppose that four control chart run rules, each with false alarm / type 1 error rate of about 0.5%, are applied to four control charts. What is the overall false alarm / type 1 error rate for the family of rules and charts?

Solution:

$$\alpha_{family} = \sum_{i=chart} \sum_{j=rule} \alpha_{ij}$$

$$\simeq 4 \times 4 \times 0.005$$

$$\simeq 0.08$$

That is, we can expect one false alarm / type 1 error to appear at, on average, in about every $1/0.08 = 12$ sampling intervals. This rate might be acceptable if we're sampling hourly; however, we must be very careful if we intend to sample more frequently or plan to use more run rules and/or keep more charts.

Process Capability

- Quality engineers use two process capability parameters:

$$c_p = \frac{USL - LSL}{6\sigma}$$

$$c_{pk} = \frac{|NSL - \mu|}{3\sigma}$$

- c_p and c_{pk} are surrogates for the fraction defective *assuming that the process is normally distributed*:

$$\begin{aligned} 1 - p &= \Phi(LSL < x < USL; \mu, \sigma) \\ &= \Phi(6c_p - 3c_{pk} < z < 3c_{pk}) \end{aligned}$$

- Small changes in the values of c_p and c_{pk} can cause huge changes in fractions defective
- There's LOTS of bad practice out there because quality engineers don't realize how much data is required to estimate c_p and c_{pk}.

Process Capability

- A large-sample confidence interval for c_p is

$$P(\widehat{c}_p(1-\delta) < c_p < \widehat{c}_p(1+\delta)) = 1-\alpha$$

  where

$$\delta = \frac{z_{\alpha/2}}{\sqrt{2n}}.$$

- The sample size to obtain confidence interval half-width δ is

$$n = \frac{1}{2}\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$

Process Capability

Example: Calculate the sample size required to estimate the process capability parameter c_p to within $\pm 10\%$ with 95% confidence.

Solution: With relative confidence interval half-width $\delta = 0.10$ and $(1-\alpha)100\% = 95\%$ confidence the required sample size is:

$$n = \frac{1}{2}\left(\frac{1.96}{0.10}\right)^2 = 193$$

Note that knowing c_p to within $\pm 10\%$ still covers a huge, possibly unacceptably wide, range of fraction defective. For example, if the 95% confidence interval for c_p is 1.5 ± 0.15 the ratio of the fraction defectives when $c_p = 1.65$ to $c_p = 1.35$ is 64!

Process Capability

- A large-sample confidence interval for c_{pk} is

$$P(\widehat{c}_{pk}(1-\delta) < c_{pk} < \widehat{c}_{pk}(1+\delta)) = 1-\alpha,$$

  where

$$\delta = z_{\alpha/2}\sqrt{\frac{1}{n}\left(\frac{1}{9\widehat{c}_{pk}^2} + \frac{1}{2}\right)}.$$

- The sample size to obtain confidence interval half-width δ is

$$n \simeq \left(\frac{z_{\alpha/2}}{\delta}\right)^2\left(\frac{1}{9\widehat{c}_{pk}^2} + \frac{1}{2}\right).$$

  When c_{pk} is very large, this reduces to the sample size required for estimating c_p.

Process Capability

Example: What sample size is required to estimate c_{pk} to within 5% of its true value with 90% confidence if $c_{pk} = 1.0$ is expected?
Solution: With $\delta = 0.05$ and $\alpha = 0.05$ the required sample size is

$$n \simeq \left(\frac{1.645}{0.05}\right)^2 \left(\frac{1}{9(1.0)^2} + \frac{1}{2}\right) = 662.$$

Tolerance Intervals

There are two tolerance interval methods available:

- Normal distribution tolerance interval
- Nonparametric tolerance interval

Normal Distribution Tolerance Interval

- If the quality characteristic is normally distributed with known mean and standard deviation then the specifiction limits would be

$$USL/LSL = \mu \pm z_{p/2}\sigma$$

  where p is the allowed fraction defective.
- When μ and σ must be estimated from sample data, it is necessary to incorporate the confidence intervals for μ and σ into the specification limit calculation. The result is the *normal distribution tolerance interval*:
  - For a two-sided specification:

$$LSL/USL = \bar{x} \pm k_2 s$$

    where $k_2 = f(p, \alpha, n)$.
  - One-sided specification:

$$LSL = \bar{x} - k_1 s$$

$$USL = \bar{x} + k_1 s$$

Tolerance Interval Factors for Normal Distributions (95% Confidence)

	One-sided (k_1)		Two-sided (k_2)	
	Yield		Yield	
n	0.99	0.999	0.99	0.999
10	3.98	5.20	4.43	5.65
15	3.52	4.61	3.88	4.95
20	3.29	4.32	3.62	4.61
25	3.16	4.14	3.46	4.41
30	3.06	4.02	3.35	4.28
40	2.94	3.87	3.21	4.10
50	2.86	3.77	3.13	3.99

Normal Distribution Tolerance Interval

Example: What are the two-sided, 99.9% yield, 95% confidence normal distribution tolerance limits for a random sample of size $n = 40$?

Solution: From the table with $n = 40$, 99.9% yield, we have $k_2 = 4.10$, so the tolerance limits are:

$$USL/LSL = \bar{x} \pm 4.1s$$

For specification limits set this way, we can be 95% confident that 99.9% of the population will fall in spec.

Example: What sample size is required to set two-sided, 99.9% yield, 95% confidence normal distribution tolerance limits at $USL/LSL = \bar{x} \pm 4s$?

Solution: From the k_2 table for 99.9% yield the required sample size is $n = 50$.

Note that this method of setting the tolerance limits does not take into account variation in future production. If typical process control methods are used to monitor the process, then shifts in the mean of about 1σ will be expected before the shift is identified and corrected, so the tolerance limits might be padded by approximately 1σ to protect against such shifts giving:

$$\begin{aligned} USL/LSL &= \bar{x} \pm (4+1)s \\ &= \bar{x} \pm 5s \end{aligned}$$

Nonparametric Tolerance Interval

- When we don't know the distribution shape we can set *nonparametric tolerance limits* equal to the minimum and maximum observed values in a sufficiently large sample to have $100(1-\alpha)\%$ confidence that the defective rate is less than $100p_U\%$; that is:

$$P(P(x_{\min} < x < x_{\max}) > (1-p_U)) = 1-\alpha$$

- The required sample size is given by

$$n \simeq \frac{\chi^2_{1-\alpha,4}}{2p_U}$$

Nonparametric Tolerance Interval

Example: What sample size is required to be 95% confident that at least 99% of a population of continuous measurement values falls within the extreme values of the sample?

Solution: With $\alpha = 0.05$ and $p_U = 0.01$, the required sample size is

$$n \simeq \frac{\chi^2_{0.95,4}}{2 \times 0.01}$$
$$\simeq 475.$$

That is, draw and measure a random sample of $n = 475$ units. Set the specification limits to

$$(LSL < x < USL) = (x_{\min} < x < x_{\max})$$

and 99% of the population should fall within those limits with 95% confidence.

Acceptance Sampling by Attributes

(From Mathews, Sample Size Calculations, p. 80) Given two points on an operating characteristic (OC) curve corresponding to an Acceptable Quality Level (AQL) condition and a Rejectable Quality Level (RQL) condition:

$$P_A(p_0 = AQL) = 1 - \alpha$$

$$P_A(p_1 = RQL) = \beta,$$

the sample size required to satisfy both conditions is

$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right)^2 .$$

Acceptance Sampling by Attributes

Example: Determine the sample size for the attributes sampling plan that will accept 95% of the lots with 0.1% defectives and reject 95% of the lots with 0.4% defectives.

Solution: The two specified points on the OC curve are $(p_0 = 0.001, P_A = 1 - \alpha = 0.95)$ and $(p_1 = 0.004, P_A = \beta = 0.005)$ so the sample size is

$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right)^2$$

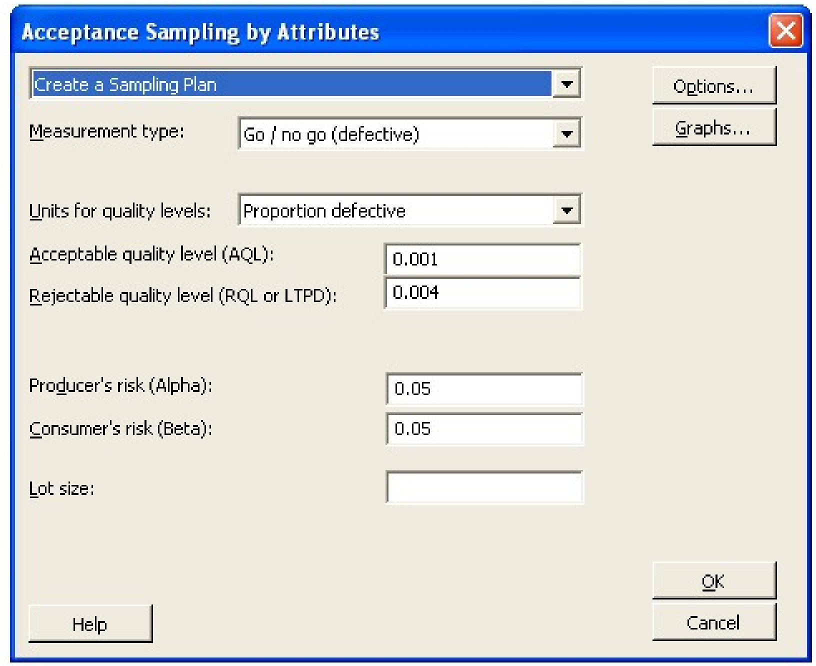
$$= \left(\frac{1.645\sqrt{0.001(0.999)} + 1.645\sqrt{0.004(0.996)}}{0.004 - 0.001} \right)^2$$

$$= 2698$$

Using MINITAB **Stat**> **Quality Tools**> **Acceptance Sampling by Attributes**

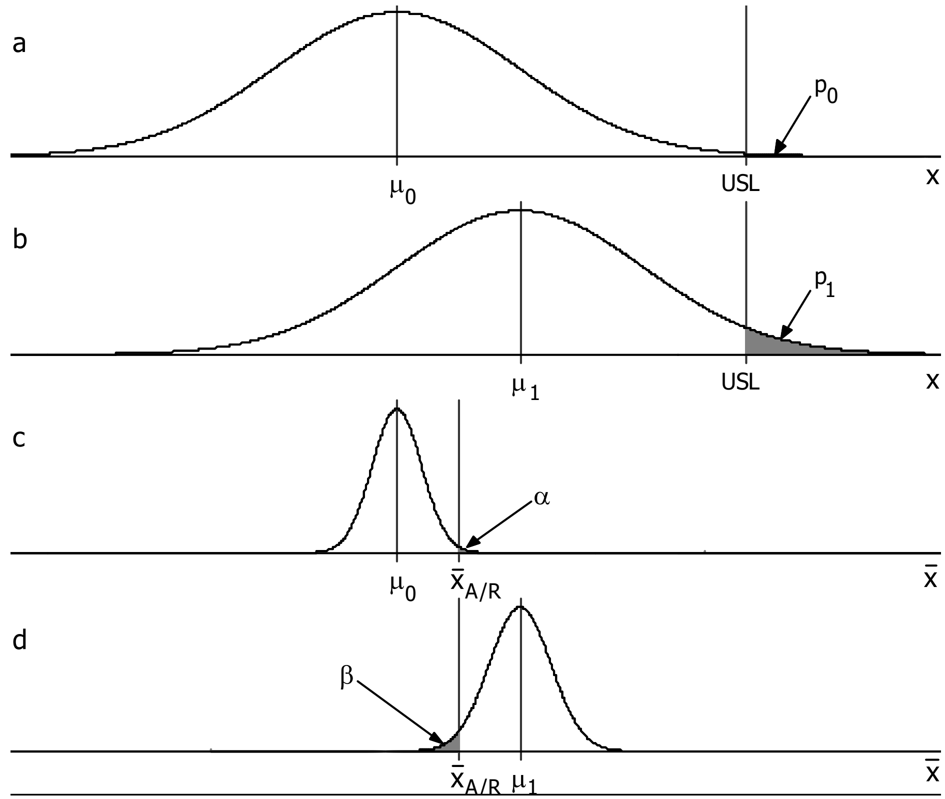
Acceptance Sampling by Attributes

```
Measurement type:  Go/no go
Lot quality in proportion defective
Use binomial distribution to calculate probability of acceptance


Acceptable Quality Level (AQL)           0.001
Producer's Risk (Alpha)                  0.05

Rejectable Quality Level (RQL or LTPD)   0.004
Consumer's Risk (Beta)                   0.05


Generated Plan(s)

Sample Size          2958
Acceptance Number       6

Accept lot if defective items in 2958 sampled <= 6;  Otherwise reject.


Proportion  Probability  Probability
 Defective    Accepting    Rejecting
     0.001        0.969        0.031
     0.004        0.050        0.950
```



Acceptance Sampling by Variables

An acceptance sampling plan for variable/measurement data can be used to control the fraction defective relative to specification limits on the variable/measurement response. The decision to accept or reject lots is based on the sample mean $\bar{x}$ and either the known value of the population standard deviation σ or the sample standard deviation s.



Acceptance Sampling by Variables

Given the AQL and RQL conditions:

$$P_A(p_0 = AQL) = 1 - \alpha$$

$$P_A(p_1 = RQL) = \beta$$

- When the population standard deviation is known the required sample size is (From Mathews, Sample Size Calculations, p. 252, Equation 10.79)

$$n = \left(\frac{z_\alpha + z_\beta}{z_{p_0} - z_{p_1}}\right)^2$$

- When the population standard deviation is unknown the required sample size is (From Schilling, Acceptance Sampling in Quality Control)

$$n = \left(1 + \frac{k^2}{2}\right)\left(\frac{z_\alpha + z_\beta}{z_{p_0} - z_{p_1}}\right)^2$$

where

$$k = \frac{z_{p_1} z_\alpha + z_{p_0} z_\beta}{z_\alpha + z_\beta}.$$

Acceptance Sampling by Variables

Example: Determine the sample size for the variables sampling plan that will accept 95% of the lots with 0.1% defectives and reject 95% of the lots with 0.4% defectives. Assume that σ is known.

Solution: The sample size is given by

$$n = \left(\frac{z_\alpha + z_\beta}{z_{p_0} - z_{p_1}} \right)^2$$

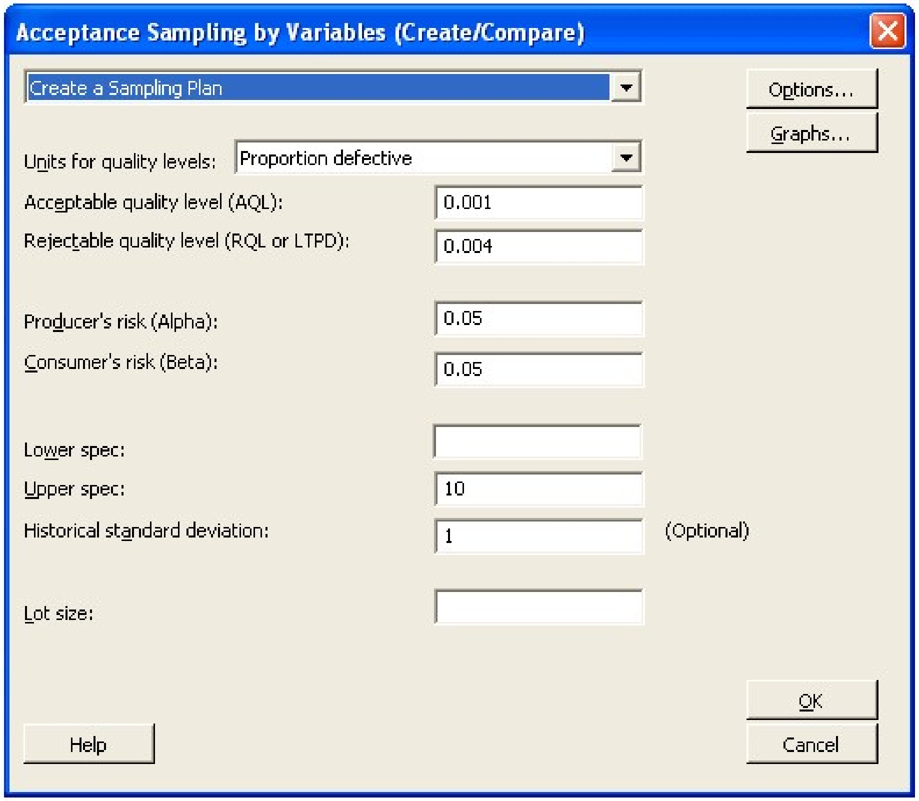
$$= \left(\frac{1.645 + 1.645}{3.09 - 2.652} \right)^2$$

$$= 57$$

Using MINITAB **Stat**> **Quality Tools**> **Acceptance Sampling by Variables**

Acceptance Sampling by Variables - Create/Compare

```
Lot quality in proportion defective


Upper Specification Limit (USL)        10
Historical Standard Deviation          1

Acceptable Quality Level (AQL)         0.001
Producer's Risk (Alpha)                0.05

Rejectable Quality Level (RQL or LTPD) 0.004
Consumer's Risk (Beta)                 0.05


Generated Plan(s)

Sample Size                  57
Critical Distance (k Value)  2.87115

Z.USL = (upper spec - mean)/historical standard deviation
Accept lot if Z.USL >= k; otherwise reject.


Proportion  Probability  Probability
 Defective    Accepting    Rejecting
     0.001        0.951        0.049
     0.004        0.049        0.951
```



Acceptance Sampling by Variables

Example: Determine the sample size for the variables sampling plan that will accept 95% of the lots with 0.1% defectives and reject 95% of the lots with 0.4% defectives. Assume that σ **is not known**.

Solution: The sample size is

$$\begin{aligned} n &= \left(\frac{z_\alpha + z_\beta}{z_{p_0} - z_{p_1}}\right)^2 \\ &= \left(\frac{1.645 + 1.645}{3.09 - 2.652}\right)^2 \\ &= 57 \end{aligned}$$

Acceptance Sampling by Variables

Example: Repeat the preceeding sample size calculation assuming that σ **is not known**.

Solution: For the σ unknown case the k factor is

$$\begin{aligned} k &= \frac{z_{p_1} z_\alpha + z_{p_0} z_\beta}{z_\alpha + z_\beta} \\ &= \frac{3.09 \times 1.645 + 2.652 \times 1.645}{1.645 + 1.645} \\ &= 2.871 \end{aligned}$$

and the sample size is

$$\begin{aligned} n &= \left(1 + \frac{k^2}{2}\right)\left(\frac{z_\alpha + z_\beta}{z_{p_0} - z_{p_1}}\right)^2 \\ &= \left(1 + \frac{2.871^2}{2}\right)\left(\frac{1.645 + 1.645}{3.09 - 2.652}\right)^2 \\ &= 289 \end{aligned}$$

Acceptance Sampling by Variables

Acceptance Sampling by Variables - Create/Compare

Lot quality in proportion defective

Method

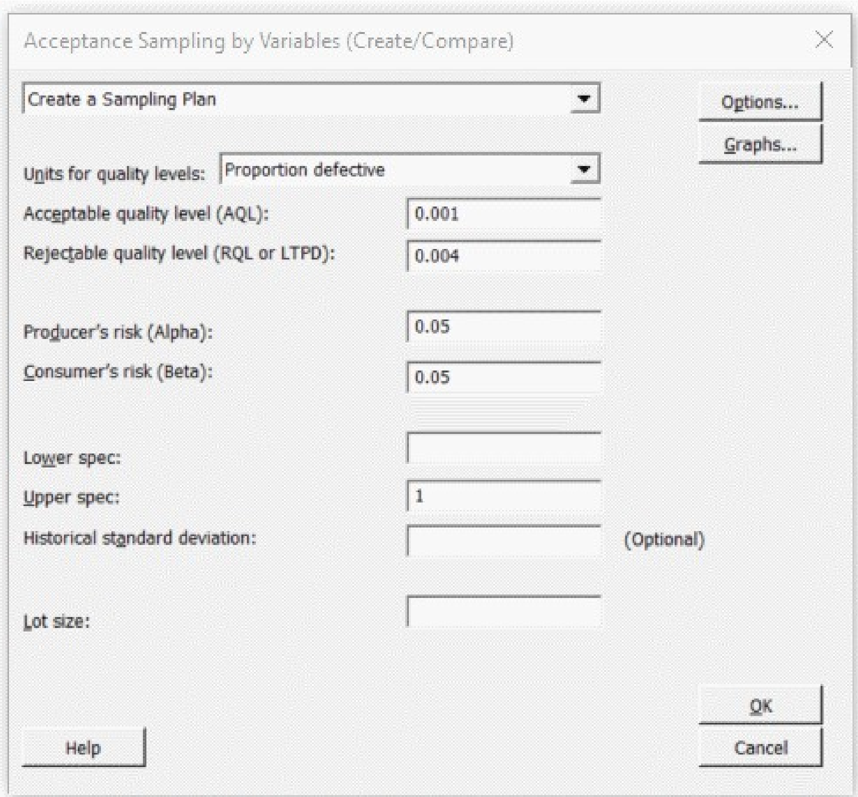
Upper Specification Limit (USL)	1
Acceptable Quality Level (AQL)	0.001
Producer's Risk (α)	0.05
Rejectable Quality Level (RQL or LTPD)	0.004
Consumer's Risk (β)	0.05

Generated Plan(s)

Sample Size	289
Critical Distance (k Value)	2.87115

Z.USL = (upper spec - mean)/standard deviation
Accept lot if Z.USL ≥ k; otherwise reject.

Proportion Defective	Probability Accepting	Probability Rejecting
0.001	0.951	0.049
0.004	0.051	0.949



Relative Efficiency of Attributes and Variables Sampling Plans

- An attribute sampling plan judges units to be in or out of specification.
- A variables sampling plan uses measurement data to assess conformance to specification.
- There are sample size calculations available for both methods. When $\alpha = \beta$ and σ for the variables sampling plan is known, the ratio of the attributes to variables sample size is approximately

$$\frac{n_{attributes}}{n_{variables}} \simeq \frac{1}{4}\left(\frac{z_{p_0} - z_{p_1}}{\sqrt{p_1} - \sqrt{p_0}}\right)^2.$$

Relative Efficiency of Attributes and Variables Sampling Plans

Example: Determine the sample size ratio for the attributes and variables inspection plans that will accept 95% of the lots with 0.1% defectives and reject 95% of the lots with 0.4% defectives. Assume that σ for the variables plan is known.

Solution: The two points on the OC curve are $(p_0 = 0.001, 1 - \alpha = 0.95)$ and $(p_1 = 0.004, \beta = 0.05)$. The ratio of the attributes- to variables-based sample sizes is approximately

$$\frac{n_{attributes}}{n_{variables}} \simeq \frac{1}{4}\left(\frac{z_{0.001} - z_{0.004}}{\sqrt{0.004} - \sqrt{0.001}}\right)^2$$

$$\simeq \frac{1}{4}\left(\frac{3.090 - 2.652}{\sqrt{0.004} - \sqrt{0.001}}\right)^2$$

$$\simeq 48$$

which is in excellent agreement with the exact ratio from the MINITAB solutions:

$$\frac{n_{attributes}}{n_{variables}} \simeq \frac{2958}{57}$$

$$\simeq 52$$

Gage Error Studies

- Quality engineers use gage error studies to validate measurement methods. (Gage error studies are analogous to the requirements of FDA, CVM64: Analytical Method Validation.)
- A typical gage error study uses several operators who measure the same units two or more times.
- The data are analyzed by random effects ANOVA and variance components analysis. Variance components are used to estimate *repeatability* or *equipment variation* (EV) and *reproducibility* or *appraiser variation* (AV).
- Acceptance criterion for EV and AV is that they must be less than 10% of the tolerance for a good measurement and less than 30% of the tolerance for a marginal measurement.

Gage Error Studies

- The confidence interval for EV is

$$P\left(0 < EV < \sqrt{\frac{df_\epsilon}{\chi^2_{\alpha,df_\epsilon}}}\,\widehat{EV}\right) = 1 - \alpha.$$

- An approximate confidence interval for AV is

$$P\left(0 < AV < \sqrt{\frac{df_O}{\chi^2_{\alpha,df_O}}}\,\widehat{AV}\right) = 1 - \alpha.$$

df	$\sqrt{df/\chi^2_{0.05}}$
1	15.95
2	4.415
3	2.920
4	2.372
6	1.915

df	$\sqrt{df/\chi^2_{0.05}}$
40	1.228
50	1.199
80	1.150
100	1.100
300	1.050

Gage Error Studies

Example: Estimate the upper confidence limits on EV/Tol and AV/Tol if a crossed GR&R study with 3 operators, 10 parts, and 2 trials delivers $\widehat{EV}/Tol = 10\%$ and $\widehat{AV}/Tol = 10\%$. If the results are unacceptable, recommend a new experiment design.

Solution: Ignoring the operator by part interaction, the ANOVA will have $df_O = 2$ and $df_\epsilon = 48$. From the table of multipliers for the upper confidence limits, $UCL_{\widehat{EV}/Tol} = 1.2 \times 0.10 = 0.12$ or 12% which is marginal and $UCL_{\widehat{AV}/Tol} = 4.4 \times 0.10 = 0.44$ or 44% which is very bad.

The problem is the low number of degrees of freedom for estimating AV which can only be resolved by using more operators. With 7 operators ($df_O = 6$) the new AV upper confidence limit would be $UCL_{\widehat{AV}/Tol} = 1.9 \times 0.10 = 0.19$ or 19% which is marginal. Using any more operators is impractical. The number of parts could be reduced to 6 or 8 to keep the total number of measurements reasonable as long as the variation in the parts is enough to challenge the operators.

Gage Error Studies

Recommendations:

- The number of parts affects EV but not AV. Use enough parts to challenge the operators.
- Use as many operators as possible — two or three are insufficient. With seven operators ($df_O = 6$), the upper confidence limit on AV will be about twice the point estimate.
- The number of trials affects EV but not AV. Two trials are usually sufficient. Three may be a waste of time.

Seminar Outline

1. Review of Fundamental Concepts
2. Means
3. Standard Deviations
4. Proportions
5. Counts
6. Linear Regression
7. Correlation
8. Designed Experiments
9. Reliability
10. Statistical Quality Control
11. **Resampling Methods**

Resampling Methods

- Monte Carlo - Resamples drawn from an assumed parametric distribution
- Bootstrap - Resamples drawn from the original sample

References

- Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences*
- Ryan (2013) *Sample Size Determination and Power*
- Mathews (2010). *Sample Size Calculations: Practical Methods for Engineers and Scientists*
- Lenth (2001). "Some Practical Guidelines for Effective Sample-Size Determination" (www.stat.uiowa.edu/techrep/tr303.pdf)
- Burdick, Borror, and Montgomery (2005). *Design and Analysis of Gage R&R Studies*
- Schilling and Neubauer (2009). *Acceptance Sampling in Quality Control*
- Meeker, Hahn, and Escobar (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*
- NCSS PASS documentation: https://www.ncss.com/software/pass/pass-documentation/