Design of Experiments

Paul G. Mathews, Mathews Malnar and Bailey, Inc.

Rev. 20230103

Course Content

What's Not Covered:

- Repeated measures
- Taguchi designs
- Optimal designs
- Mixture experiments
- Split-plot designs
- Analysis of qualitative (i.e. binary, nominal, and ordinal) responses

DOE References:

- Montgomery, *Design and Analysis of Experiments*, Wiley.
- Box, Hunter, and Hunter, *Statistics for Experimenters*, Wiley.
- Hicks, *Fundamental Concepts in the Design of Experiments*, Saunders College Publishing.
- Mathews, *Design of Experiments with MINITAB*, ASQ Quality Press.
- Bhote and Bhote, *World Class Quality: Using Design of Experiments to Make It Happen*, AMACOM.
- Neter, Kutner, Nachtscheim, and Wasserman, *Applied Linear Statistical Models*, McGraw-Hill.

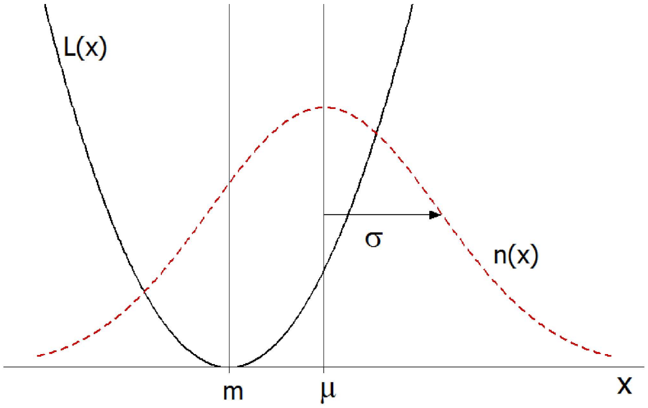
Definitions

- What is an *experiment*?
  - An activity that includes collection and analysis of data and interpretation of the results for the purpose of managing a process.
- The simplest experiment:
  - Collect a representative sample from a single stable process
  - Measure the sample
  - Calculate sample statistics (point estimates) for the mean and standard deviation
  - Calculate relevant confidence intervals or perform hypothesis tests
  - Check distribution shape
  - Interprete the results
- What is a *designed* experiment?
  - A carefully structured experiment with highly desireable mathematical and statistical properties designed to answer specific research questions about the values of a population's parameters and/or distribution shape.

Motivations for DOE

Recall Taguchi's Loss Function:

$$\bar{L} = k\left((\mu - m)^2 + \sigma^2\right)$$

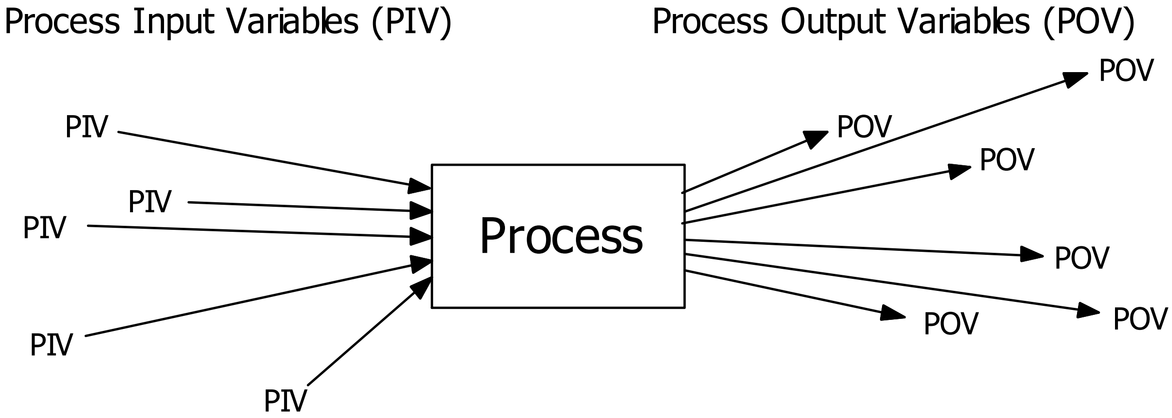


Motivations for DOE

- The purpose of DOE is to determine how a response (y) depends on one or more input variables or predictors (x_i) so that future values of the response can be predicted from the input variables.
- DOE methods are necessary because the one variable at a time (OVAT) method (that is, changing one variable at a time while holding all the others constant) cannot account for interactions between variables.
- DOE requires you to change how you do your work but it does not increase the amount of work you have to do. DOE allows you to learn more about your processes while doing the same or even less work.
- DOE allows you to:
  - Build a mathematical model for a response as a function of the input variables.
  - Select input variable levels that optimize the response (e.g. minimizing, maximizing, or hitting a target).
  - Screen many input variables for the most important ones.
  - Eliminate insignificant variables that are distracting your operators.
  - Identify and manage the interactions between variables that are preventing you from optimizing your design or process or that are confusing your operators.
  - Predict how manufacturing variability in the input variables induces variation in the response.
  - Reduce variation in the response by identifying and controlling the input variables are contributing the most to it.

Chapter 1: Graphical Presentation of Data

- Types of data
  - Attribute, categorical, or qualitative data, e.g. types of fruit
  - Variable, measurement, or quantitative data, e.g. lengths measured in millimeters
- Types of variables: Use an Input-Process-Output (IPO) diagram to document all of your process's process input and process output variables:



- Always graph the data!
  - Bar charts
  - Histograms
  - Dotplots
  - Stem-and-leaf plots
  - Scatter plots
  - Multi-vari charts
  - Probability plots

Chapter 2: Descriptive Statistics

- What to look for when you look at a histogram, dotplot, ... :
  - Location or central tendency
  - Variation, dispersion, scatter, noise
  - Distribution shape, e.g. bell-shaped, symmetric or asymmetric (skewed), etc.
  - Outliers
- Parameters and statistics
  - A parameter is a measure of location or variation of a population.
  - A statistic is a measure of location or variation of a sample.
  - If the sample is representative of the population, then a sample statistic might be a good estimate of a population parameter.
- Measures of location:
  - Population mean (μ)
  - Sample median $(\tilde{x})$ - middle value in the data set when the observations are ordered from smallest to largest
  - Sample mean $(\bar{x})$:

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

  - If the sample is representative of its population, then the sample mean $(\bar{x})$ might be a good estimate of the population mean (μ).
- Measures of variation:
  - Population standard deviation (σ)
  - Sample range
    - Difference between the maximum and minimum values in a sample: $R = max(x_1,x_2,\ldots) - min(x_1,x_2,\ldots)$
    - Can be used to estimate the population standard deviation:

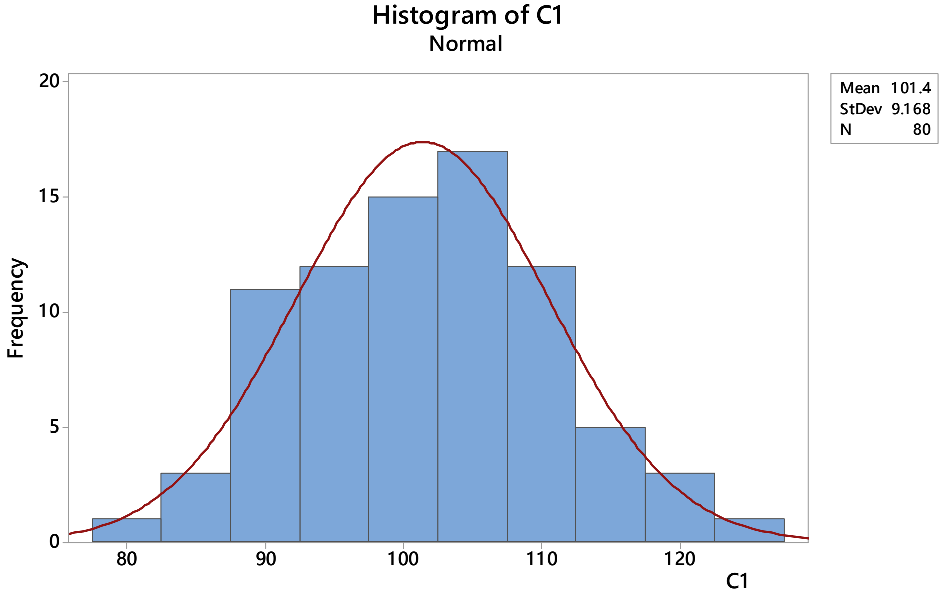
$$\sigma \simeq R/d_2$$

  - Sample standard deviation (s):

$$s = \sqrt{\frac{\sum \epsilon_i^2}{df_\epsilon}}$$

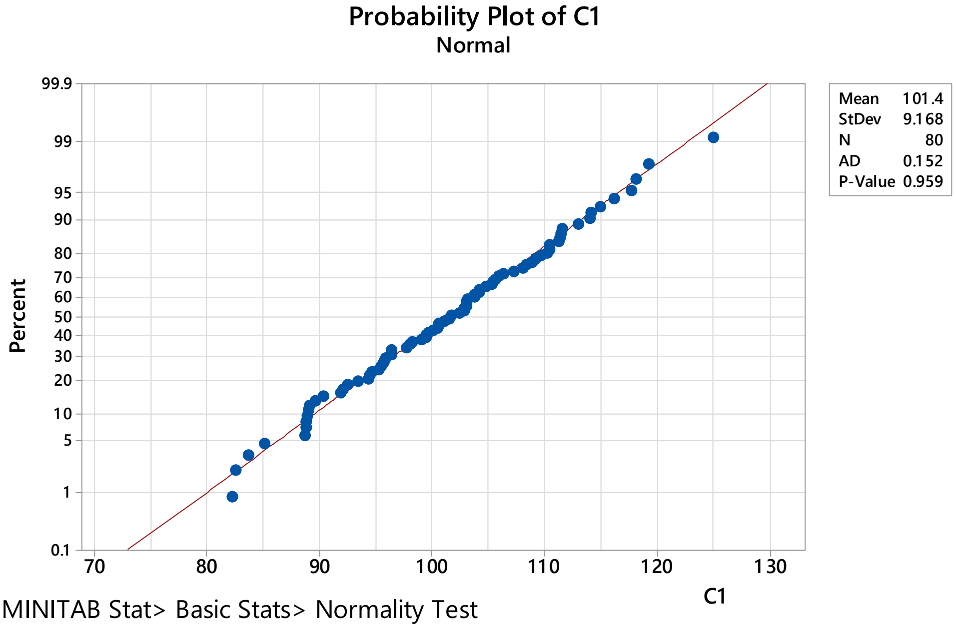
    where $\epsilon_i = x_i - \bar{x}$ and $df_\epsilon = n - 1$.
  - If the sample is representative of its population, then the sample standard deviation (s) might be a good estimate of the population standard deviation (σ).
  - Variance (s^2 or σ^2)
    - The square of the standard deviation is called the variance.
    - The variance is the fundamental measure of variation.
      - Variances can be added and subtracted from each other.
      - Ratios of variances have meaning.

- Distribution shape:
  - The most common distribution that we deal with in introductory DOE is the normal distribution, aka, the bell curve, the error function, the gaussion distribution
  - Whether or not a sample appears to follow a normal distribution is often judged by inspecting a histogram with a superimposed normal curve.



MINITAB Graph> Histogram> With fit

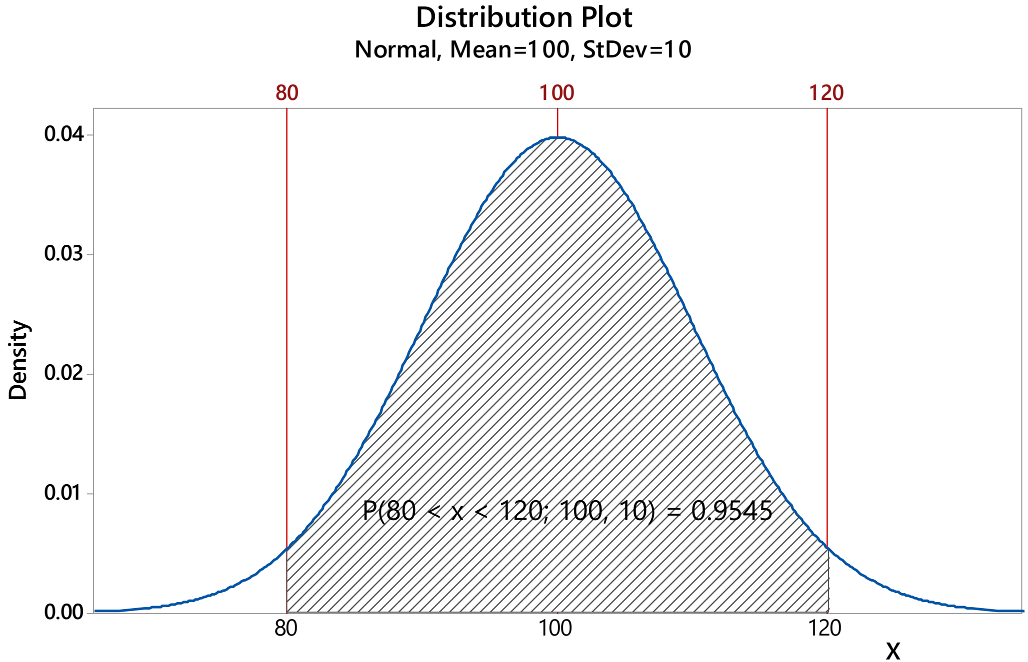
  - Normal Probability Plots
    - The much-preferred method for judging normality is using a normal probability plot.
    - A normal plot is a mathematical transformation of a histogram and its superimposed bell curve.
    - The raw data values (x) are plotted on one axis and the expected positions of those data values under the assumption of a normal distribution $(E(x|x{\sim}\Phi))$ are plotted on the other axis.
    - If the distribution is normal then the plotted points will fall along a straight line.



MINITAB Stat> Basic Stats> Normality Test

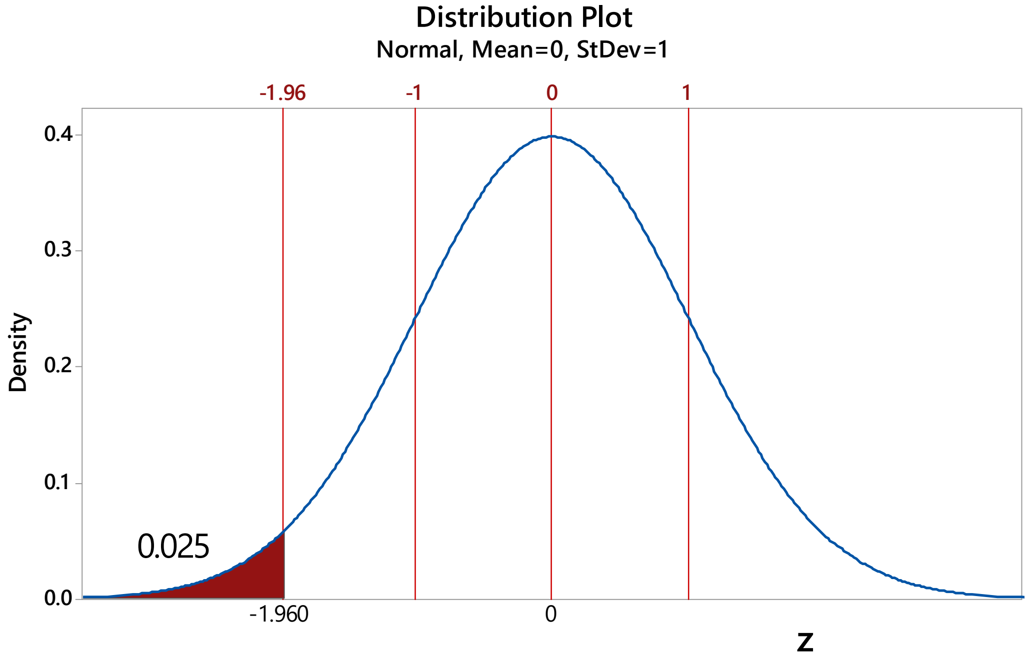
Working With the Normal Distribution

- The normal distribution is normalized so that the area under the curve is exactly 1.0. Then a vertical slice of the normal distribution can be interpreted as the probability of the variable taking on the slice's range of values.



Graph> Probability Distribution Plot> View Probability

- The *standard normal curve*:
  - Has $\mu = 0$ and $\sigma = 1$.
  - Is the distribution that is tabulated in the tables in the backs of statistics textbooks. e.g. Table A.2 on p. 478 of DOE with MINITAB



Working With the Normal Distribution

- Solving problems stated in measurement (x) units requires that we be able to transform from those units and standard (z) units and back again.

$$z = \frac{(x-\mu)}{\sigma}$$

$$x = \mu + z\sigma$$

Example: Find the fraction defective produced by a process to specification $USL/LSL = 0.440 \pm 0.020$ inches if the mean of the process is $\mu = 0.445$ inches and the standard deviation is $\sigma = 0.010$ inches. Assume that the distribution is normal.

Solution: We need to find:

$$\Phi(0.420 < x < 0.460; \mu = 0.445, \sigma = 0.010)$$

If we apply the standardizing transformation to the LSL:

$$\begin{aligned} z_{LSL} &= \tfrac{LSL-\mu}{\sigma} \\ &= \tfrac{0.420-0.445}{0.010} = -2.50 \end{aligned}$$

Similarly the z value of the USL is $z_{USL} = \frac{0.460-0.445}{0.010} = 1.50$.

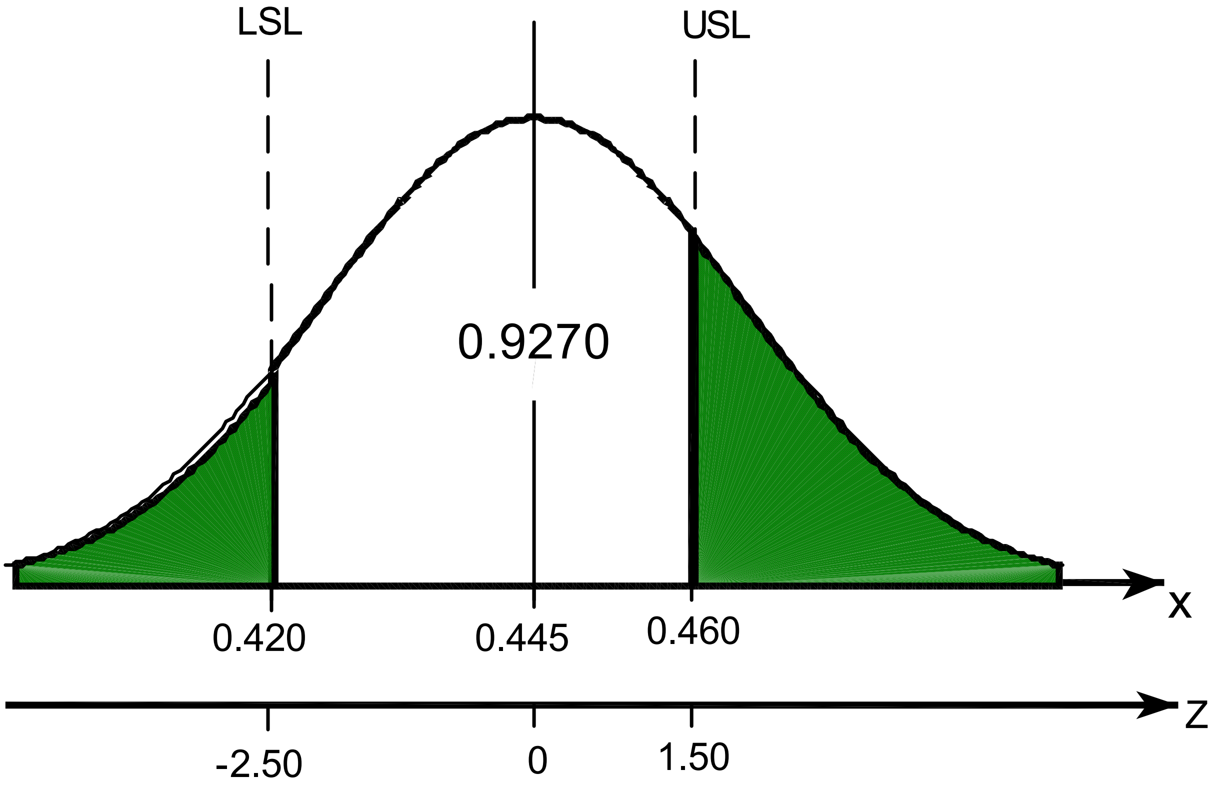
Now our interval on x:

$$\Phi(0.420 < x < 0.460; 0.445, 0.010)$$

becomes an interval on z that we can evaluate from the normal tables:

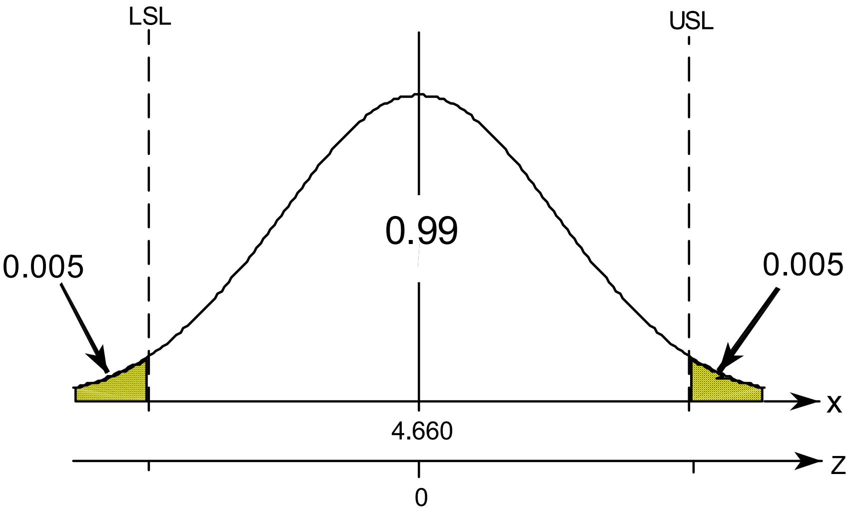
$$\begin{aligned} \Phi(-2.50 < z < 1.50) &= 0.9332 - 0.0062 \\ &= 0.9270 = 1 - 0.0730 \end{aligned}$$

This means that 92.7% of the product is in spec and 7.3% of the product is out of spec.



Example: Determine a two-sided specification for a process that has $\mu = 4.660$ and $\sigma = 0.008$ if the specification must contain 99% of the population. Assume that the distribution is normal.

Solution:



If 99% of the product must be in the symmetric two-sided specification then there will be 0.5% of the product out of spec on the high and low ends of the distribution. Since $z_{0.005} = 2.575$ the required specification is:

$$\Phi(LSL < x < USL; 4.660, 0.008) = 0.99$$

where

$$\begin{aligned} LSL &= \mu - z_{0.005}\sigma \\ &= 4.660 - 2.575 \times 0.008 \\ &= 4.639 \end{aligned}$$

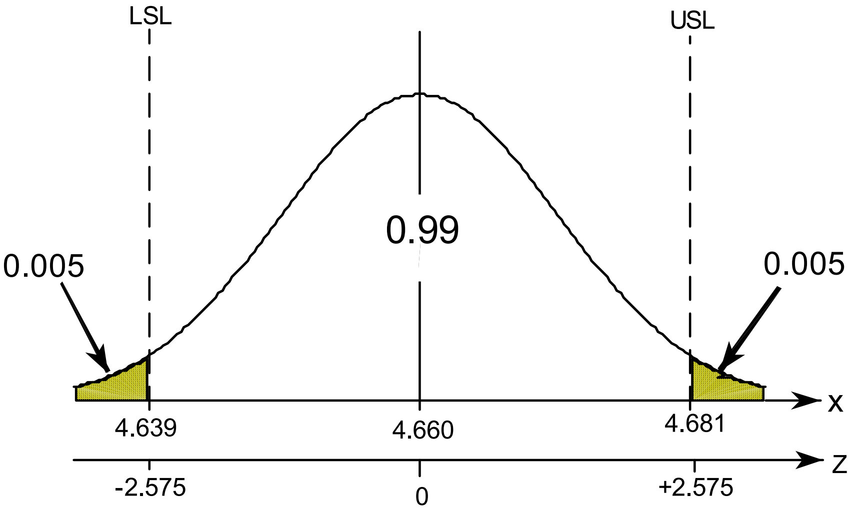
and

$$\begin{aligned} USL &= \mu + z_{0.005}\sigma \\ &= 4.660 + 2.575 \times 0.008 \\ &= 4.681 \end{aligned}$$

Finally we have:

$$\Phi(4.639 < x < 4.681; 4.660, 0.008) = 0.99$$

so our spec of $USL/LSL = 4.681/4.639$ will contain 99% of the population.



Counting

- Multiplication of choices
- Factorials
- Permutations
- Combinations

Counting: Multiplication of Choices

If a series of k decisions must be made and the first can be made in n_1 ways, the second in n_2 ways, and so on, then the total number of different ways that all k decisions can be made, n_{total}, is:

$$n_{total} = n_1 n_2 \cdots n_k$$

Example: If an arc lamp experiment is going to be constructed and there are five arctube designs, three mount designs, two bulb types, and four bases, how many unique configurations can be constructed?

Solution: Since $n_{total} = 5 \times 3 \times 2 \times 4 = 120$ there are 120 unique lamp configurations. This experiment design is called a *full factorial design*.

Counting: Factorials

If there are n distinct objects in a set and all n of them must be picked then the total number of different ways they can be picked is:

$$\text{Number of ways} = n(n-1)(n-2)(n-3)\cdots(3)(2)(1) = n!$$

where ! indicates the factorial operation.

Counting: Permutations

- If there are n distinct objects in a set and r of them are to be picked where *the order in which they are picked is important*, then there are ${}_nP_r$ ways to make the selections where:

$$\begin{aligned} {}_nP_r &= n(n-1)(n-2)\ldots(n-r+1) \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

- Derivation:

$$n! = \overbrace{\underbrace{n(n-1)(n-2)\ldots(n-r+1)}_{{}_nP_r}}\underbrace{(n-r)\ldots 3 \cdot 2 \cdot 1}_{(n-r)!}$$

$${}_nP_r = \frac{n!}{(n-r)!}$$

Example: How many different ways can a salesman fly to 5 different cities if there are 8 cities in his territory?

Solution: The number of five-city flight plans is:

$$\begin{aligned} {}_8P_5 &= \frac{8!}{(8-5)!} \\ &= \frac{8!}{3!} \\ &= \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3!}{3!} \\ &= 6720 \end{aligned}$$

Counting: Combinations

- In many situations we do not care about the order that the objects are obtained, only how many different sets of selections are possible. In these cases the permutation over-counts by a factor of ${}_rP_r$.
- If there are n objects in a set and r of them are to be picked and the order in which the picked objects are received is not important then there are ${}_nC_r$ ways to make the selections where:

$$ {}_nC_r = \binom{n}{r} = \frac{{}_nP_r}{{}_rP_r} = \frac{n!}{r!(n-r)!} $$

Example (**revisiting the air-travelling salesman**): How many different sets of five cities can the salesman visit if there are 8 cities in his territory?

Solution: The number of sets of five cities he has to select from is:

$$ \begin{aligned} \binom{8}{5} &= \frac{8!}{5!(8-5)!} \\ &= \frac{8\times7\times6\times5!}{5!3!} \\ &= 56 \end{aligned} $$

Example: Product supplied from five different vendors is to be tested and compared for differences in location. If each vendor's mean is compared to every other vendor's mean then how many tests have to be performed?

Solution:

$$ \binom{5}{2} = \frac{5!}{2!3!} = \frac{5\times4\times3!}{2!3!} = 10 $$

If the numbers 1 through 5 are used to indicate the five vendors, then the two-vendor *multiple comparisons tests* that must be performed are: 12, 13, 14, 15, 23, 24, 25, 34, 35, 45.

Example: An experiment with six variables is to be performed. If we are concerned about the possibility of interactions between variables, then how many two-factor and three-factor interactions are there?

Solution:

$$ \binom{6}{2} = \frac{6!}{2!4!} = \frac{6\times5\times4!}{2!4!} = 15 $$

$$ \binom{6}{3} = \frac{6!}{3!3!} = \frac{6\times5\times4\times3!}{3!3!} = 20 $$

The two-factor interactions are: 12, 13, 14, 15, 16, 23, 24 25, 26, 34, 35, 36, 45, 46, 56 and the three-factor interactions are: 123, 124, 125, 126, 134, 135, 136, 145, 146, 156, 234, 235, 236, 245, 246, 256, 345, 346, 356, 456.

Example: A person is on 10 different medications. In addition to the good and bad effects of each medication there is a risk of interactions between drugs. How many different two drug interactions must the doctor be aware of in treating this person? Three drug interactions?

Solution: There are $\binom{10}{2} = 45$ possible two drug interactions and $\binom{10}{3} = 120$ possible three drug interactions.

Chapter 3: Inferential Statistics

Analysis of Experimental Data

- Data from experiments are analyzed for the values of distribution parameters (e.g. mean and standard deviation) and distribution shape (e.g. normal).
- Point estimates for the distribution parameters are insufficient; hypothesis tests and confidence intervals that make probabalistic statements about their values are necessary.

Review: Limits on a Population

Example: A population (x) has $\mu_x = 320$, $\sigma_x = 20$, and is normally distributed. Find a symmetric interval on x that contains 95% of the population.

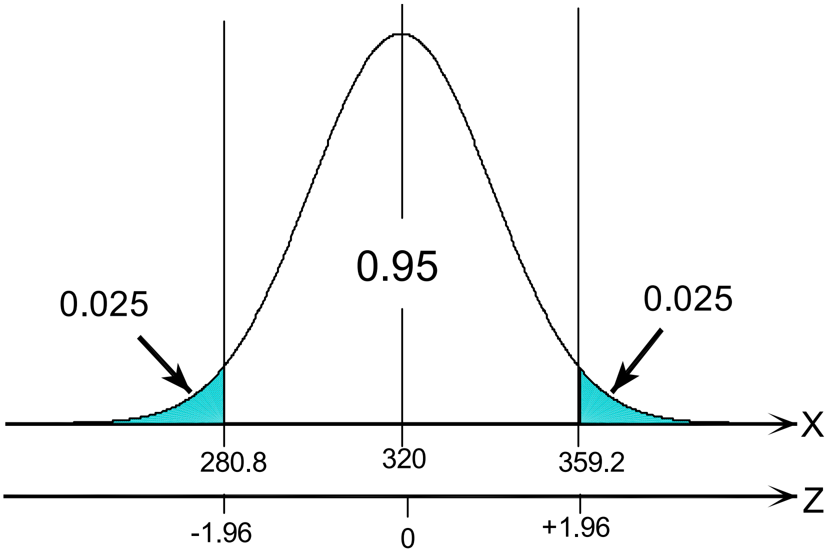
Solution: The required interval is given by:

$$\Phi(\mu_x - z_{\alpha/2}\sigma_x < x < \mu_x + z_{\alpha/2}\sigma_x) = 1 - \alpha$$

Since $1 - \alpha = 0.95$ we have $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$. The required interval becomes:

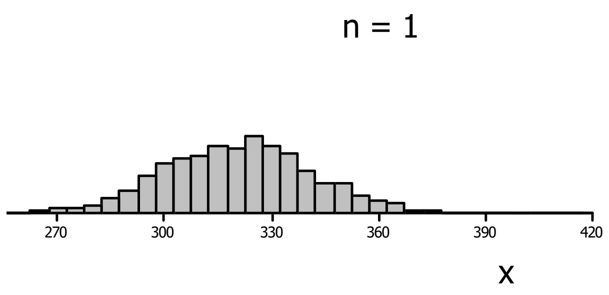
$$\Phi(320 - 1.96(20) < x < 320 + 1.96(20)) = 1 - 0.05$$

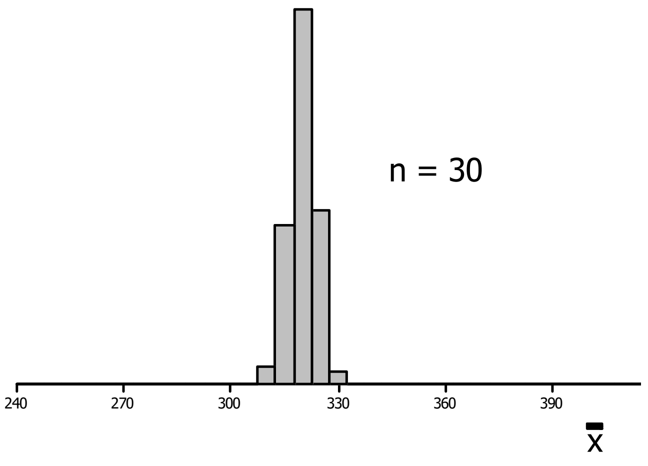
$$\Phi(280.8 < x < 359.2) = 0.95$$



Gedanken Experiment

Suppose that we compare the histogram of the measurements from 1000 samples taken from a normal distribution with $\mu = 320$ and $\sigma = 20$ to the histogram of the sample means for samples of size $n = 30$ taken from the same population:





The Central Limit Theorem

The distribution of sample means ($\bar{x}$) for samples of size n is normal (Φ) with mean:

$$\mu_{\bar{x}} = \mu_x$$

and standard deviation:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

if the following conditions are met:

1. The population standard deviation σ_x is known or the sample size is very large $(n \geq 30)$ so that σ_x can be approximated with the sample standard deviation s.
2. The distribution of the population (x) is normal.

The central limit theorem is very robust to deviations from these conditions so the scope of its applications is very broad.

Using the Central Limit Theorem

An immediate application of the Central Limit Theorem is for the calculation of an interval that contains a specified fraction of the expected sample means. Given μ_x, σ_x, n, and α the interval that contains $(1-\alpha)100\%$ of the expected sample means is:

$$\Phi(\mu_x - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu_x + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

where

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Limits on Sample Means

Example: Samples of size $n = 30$ are drawn from a population that has $\mu_x = 320$ and $\sigma_x = 20$. Find a symmetric interval that contains 95% of the sample means.

Solution: Since the sample size is large the Central Limit Theorem is valid. The required interval for $\bar{x}$s is given by:

$$\Phi(\mu_x - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu_x + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

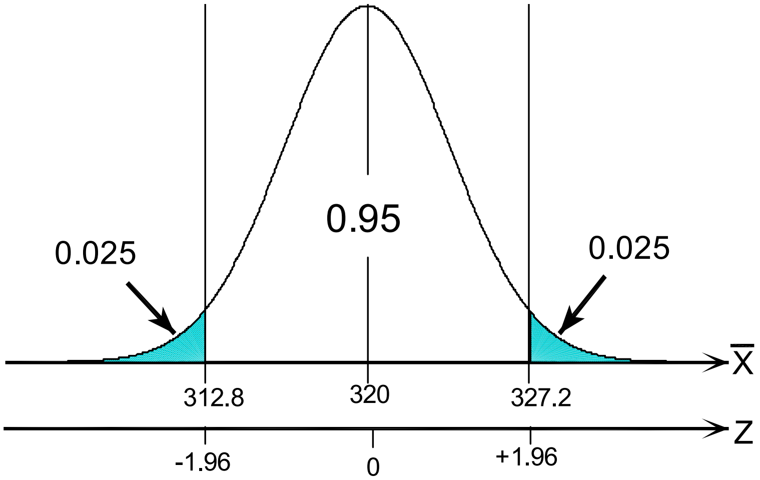
Since $1 - \alpha = 0.95$ we have $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$. The standard deviation of the $\bar{x}$s is

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{20}{\sqrt{30}} = 3.65$$

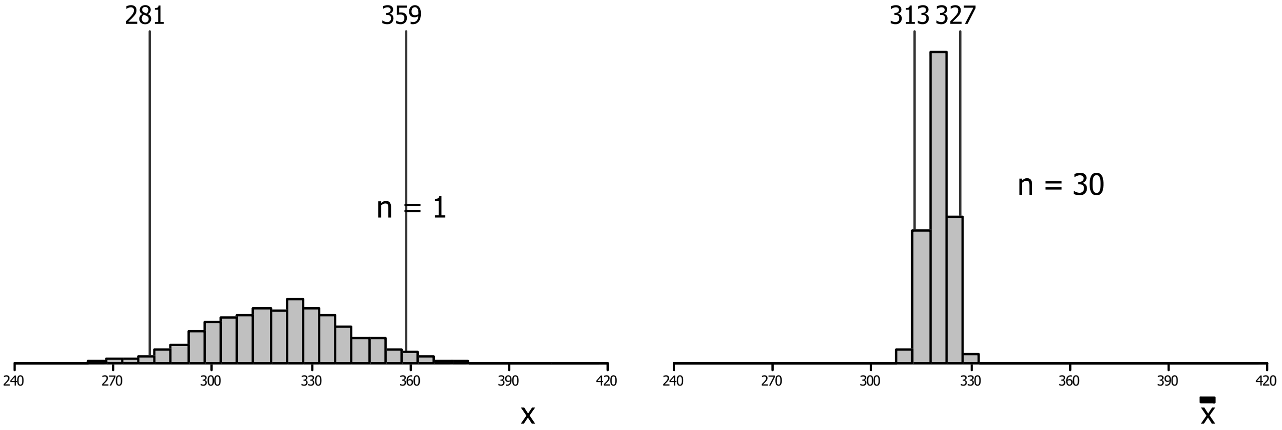
The required interval becomes:

$$\Phi(320 - 1.96(3.65) < \bar{x} < 320 + 1.96(3.65)) = 1 - 0.05$$

$$\Phi(312.8 < \bar{x} < 327.2) = 0.95$$



Comparing the Intervals



Confidence Interval for the Population Mean

The Central Limit Theorem gives us:

$$\Phi(\mu_x - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu_x + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

The random variable $\bar{x}$ is bounded on the lower and upper sides in two inequalities:

$$\mu_x - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} \quad \text{and} \quad \bar{x} < \mu_x + z_{\alpha/2}\sigma_{\bar{x}}$$

If we solve these inequalities for μ_x we obtain:

$$\mu_x < \bar{x} + z_{\alpha/2}\sigma_{\bar{x}} \quad \text{and} \quad \bar{x} - z_{\alpha/2}\sigma_{\bar{x}} < \mu_x$$

Now, if we put these two inequalities back together:

$$\Phi(\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} < \mu_x < \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

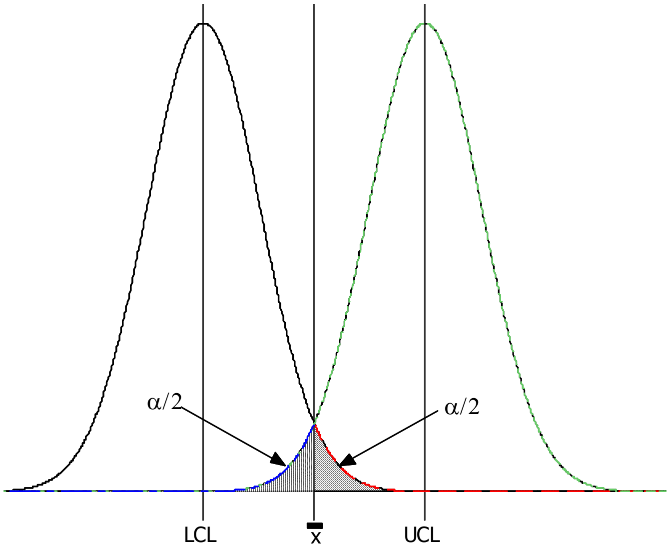
which is the two sided $(1-\alpha)100\%$ confidence interval for the unknown population mean μ_x based on a sample which has sample mean $\bar{x}$.

Graphical Interpretation

The upper and lower confidence limits given by:

$$UCL/LCL = \bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$$

represent the extreme high and low values of μ_x that could be expected to deliver the experimental $\bar{x}$ value.



Confidence Interval Example

Example: Construct a two-sided 95% confidence interval for the true population mean based on a sample of size $n = 30$ which yields $\bar{x} = 290$. The population standard deviation is $\sigma = 20$ and the distribution of the xs is normal.

Solution: Since the Central Limit Theorem is satisfied (distribution of x is normal and σ_x is known) the confidence interval is given by:

$$\Phi(\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} < \mu_x < \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

Since $\alpha = 0.05$ we have $z_{\alpha/2} = z_{0.025} = 1.96$ so:

$$\Phi\left(290 - 1.96\left(\frac{20}{\sqrt{30}}\right) < \mu_x < 290 + 1.96\left(\frac{20}{\sqrt{30}}\right)\right) = 1 - 0.05$$

The required confidence interval is:

$$\Phi(282.8 < \mu_x < 297.2) = 0.95$$

That is, we can be 95% confident that the true but unknown value of the population mean lies between 282.8 and 297.2.

Confidence Interval Interpretation

- A two-sided confidence interval for the mean has the form
$$P(LCL < \mu < UCL) = 1 - \alpha$$
- The interval $LCL < \mu < UCL$ indicates the range of possible μ values that are statistically consistent with the observed value of $\bar{x}$.
- If the confidence interval is sufficiently narrow then the interval $LCL < \mu < UCL$ will indicate a single action. Take it.
- If the confidence interval is too wide then the interval will indicate two or more actions. More data will be required.
- Ask yourself:
  - What action would I take if $\mu = LCL$?
  - What action would I take if $\mu = UCL$?
  - If the two actions are the same then take the indicated action.
  - If the two actions are different then the confidence interval is too wide. When in doubt, take more data.

Hypothesis Tests

Definition: A hypothesis test is a statistically based way of deciding which of two complementary statements about a population parameter or distribution is true on the basis of sample data. The two statements are called the null hypothesis (H_0) and the alternative hypothesis (H_A).

Hypothesis Tests

- **One population**:
  - $H_0 : \mu = 320$ versus $H_A : \mu \neq 320$ (two-tailed test)
  - $H_0 : \mu = 320$ versus $H_A : \mu < 320$ (one- / left-tailed test)
  - $H_0 : \mu = 320$ versus $H_A : \mu > 320$ (one- / right-tailed test)
  - $H_0 : \sigma = 20$ versus $H_A : \sigma \neq 20$
  - $H_0 : \sigma = 20$ versus $H_A : \sigma < 20$
  - $H_0 : \sigma = 20$ versus $H_A : \sigma > 20$
  - $H_0 : p = p_0$ versus $H_A : p \neq p_0$
  - $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda \neq \lambda_0$
  - H_0 :The distribution of x is Φ versus H_A :The distribution of x is not Φ
  - H_0 :The distribution of s^2 is χ^2 versus H_A :The distribution of s^2 is not χ^2
- **Two populations**:
  - $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$
  - $H_0 : \sigma_1 = \sigma_2$ versus $H_A : \sigma_1 \neq \sigma_2$
  - $H_0 : p_1 = p_2$ versus $H_A : p_1 \neq p_2$
  - $H_0 : \lambda_1 = \lambda_2$ versus $H_A : \lambda_1 \neq \lambda_2$
  - H_0 : *The distribution shape of* x_1 *is the same as the distribution shape of* x_2 versus H_A : *The distribution shape of* x_1 *is NOT the same as the distribution shape of* x_2.
- **Many populations**:
  - $H_0 : \mu_1 = \mu_2 = \cdots$ versus $H_A : \mu_i \neq \mu_j$ for at least one i,j pair
  - $H_0 : \sigma_1 = \sigma_2 = \cdots$ versus $H_A : \sigma_i \neq \sigma_j$ for at least one i,j pair
  - $H_0 : p_1 = p_2 = \cdots$ versus $H_A : p_i \neq p_j$ for at least one i,j pair
  - $H_0 : \lambda_1 = \lambda_2 = \cdots$ versus $H_A : \lambda_i \neq \lambda_j$ for at least one i,j pair

Which Test?

- What type of data?
  - Measurement/variable
  - Attribute
    - Binary/dichotomous, e.g. defectives
    - Count, e.g. defects
- How many populations?
- What population characteristic?
  - Location
  - Variation
  - Distribution Shape
  - Other
- Exact or approximate method?
- See QES Appendix B: Hypothesis Test Matrix

Understanding Hypotheses

- Statistical hypotheses have two forms, one stated mathematically and the other stated in the language of the context. For example, in SPC the hypothesis $H_0 : \mu = 25$ corresponds to the statement *the process is in control*.
- Sagan's Rule: *To test the hypotheses*

  H_o*: Something ordinary happens*

  versus

  H_A*: Something extraordinary happens,*

  the extraordinary claim requires extraordinary evidence.
- In quality engineering, sometimes the hypotheses are determined by historical choice:
  - SPC: H_0*: the process is in control* versus H_A*: the process is out of control*.
  - Acceptance sampling: H_0*: the lot is good* versus H_A*: the lot is bad*.

General Hypothesis Testing Procedure

1. Formulate the null (H_0) and alternative hypotheses (H_A). Put the desired conclusion in H_A.
2. Specify the significance level α (the risk of a Type 1 error).
3. Construct accept and reject criteria for the hypotheses based on the sampling distribution of an appropriate test statistic at the required significance level.
4. Collect the data and calculate the value of the test statistic.
5. Compare the test statistic to the acceptance interval and decide whether to accept or reject H_0. In practice, we never accept H_0. We either reject H_0 and accept H_A or we say that the test is inconclusive.

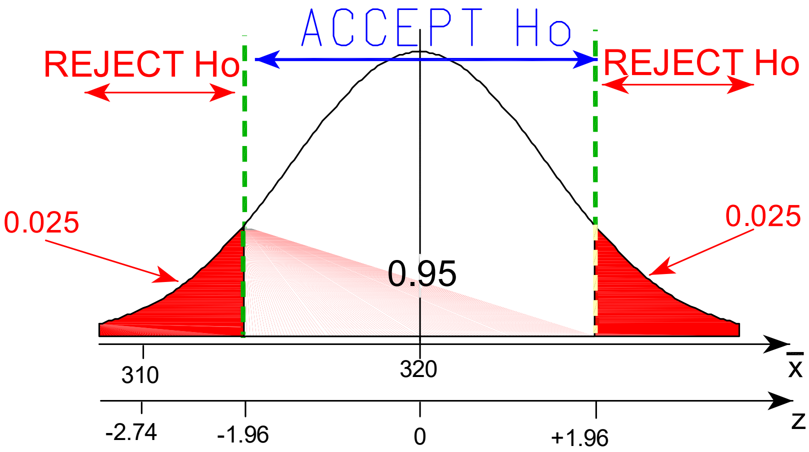
Hypothesis Test Example

Example A: Test the hypotheses $H_0 : \mu = 320$ vs. $H_A : \mu \neq 320$ on the basis of a sample of size $n = 30$ taken from a normal population with standard deviation $\sigma = 20$ which yields $\bar{x} = 310$. Use the 5% significance level.

Solution: The two hypotheses are already given to us. The appropriate statistic to test them is $\bar{x}$. If $\bar{x}$ falls very close to 320 then we will accept H_0, otherwise we will reject it. The Central Limit Theorem describes the distribution of the $\bar{x}$s and with $\alpha = 0.05$ we have a critical z value of $z_{0.025} = 1.96$. This means that we will accept H_0 if the test statistic falls in the interval $-1.96 < z < +1.96$. The z value that corresponds to $\bar{x}$ is given by:

$$
\begin{aligned}
z &= \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \\
&= \frac{310 - 320}{20/\sqrt{30}} \\
&= -2.74
\end{aligned}
$$

Since $z = -2.74$ falls outside the acceptance interval $\bar{x}$ must be significantly different from the hypothesized mean of $H_0 : \mu = 320$ so we must reject H_0 in favor of $H_A : \mu \neq 320$.



Relationship Between Confidence Intervals and Hypothesis Tests

- The confidence interval and hypothesis test provide different ways of performing the same analysis but they both offer unique features that prohibit the exclusive use of one method or the other.
- The confidence interval for the mean is centered on the sample mean:

  $$UCL/LCL = \bar{x} \pm \delta$$

  where the confidence interval half-width is

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

- The accept/reject decision limits for the hypothesis test are centered on μ_0:

  $$UDL/LDL = \mu_0 \pm \delta$$

  where δ has the same value as the confidence interval half-width.
- The confidence interval is the set of all possible values of μ_0 for which we would accept H_0, so …
- If μ_0 falls inside of the confidence limits then we accept $H_0 : \mu = \mu_0$ and if μ_0 falls outside of the confidence limits then we reject H_0.

Example: Construct the confidence interval for the population mean in Example A and use it to test the hypotheses $H_0 : \mu = 320$ vs. $H_A : \mu \neq 320$.

Solution: The confidence interval is

$$\Phi\left(310 - 1.96\left(\frac{20}{\sqrt{30}}\right) < \mu_x < 310 + 1.96\left(\frac{20}{\sqrt{30}}\right)\right) = 0.95$$

$$\Phi(302.8 < \mu_x < 317.2) = 0.95$$

The confidence interval does NOT contain $\mu = 320$ so we must reject $H_0 : \mu = 320$ in favor of $H_A : \mu \neq 320$.

Errors in Hypothesis Testing

There are two kinds of errors that can occur in hypothesis testing:

1. Type 1 Error: We reject the null hypothesis when it is really true.
2. Type 2 Error: We accept the null hypothesis when it is really false.

These errors and the situations in which correct decisions are made are summarized in the following table:

The truth is:→	H_0 is true	H_0 is false
The test says accept H_0	Correct Decision	Type 2 Error
The test says reject H_0	Type 1 Error	Correct Decision

Errors in the Legal System

- Hypotheses:
  - H_0 :The defendant is not guilty
  - H_A :The defendant is guilty
- Quiz: Was the correct decision made and, if not, what type of error occurred?
  - A not guilty verdict for an innocent person.
  - A guilty verdict for an innocent person.
  - A guilty verdict for a guilty person.
  - A not guilty verdict for a guilty person.

Understanding Type 1 and Type 2 Errors

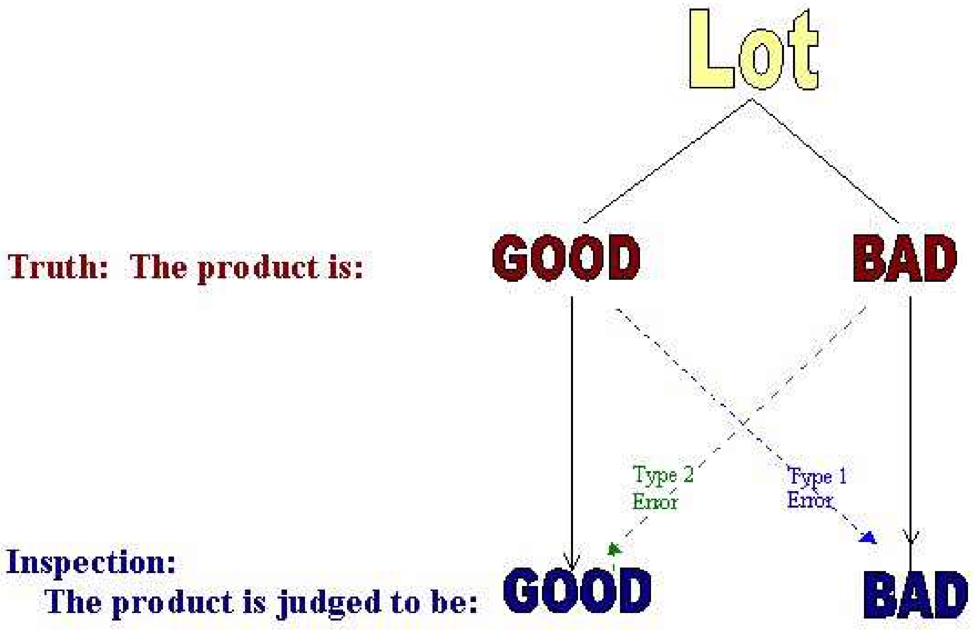
In a final inspection operation just before shipping to the customer:

- If truly good material is tested and the test returns an erroneous *Reject* H_0*: the material is bad* result then a Type 1 error has occurred. This compromises the manufacturer's position (he cannot sell this good material) so the risk of committing a Type 1 error is often called the *manufacturer's risk*.
- If truly bad material is tested and the test returns an erroneous *Accept* H_0*: the material is good* result then a Type 2 error has occurred. This compromises the consumer's position (he has just approved the use of bad material) so the risk of committing a Type 2 error is often called the *consumer's risk*.

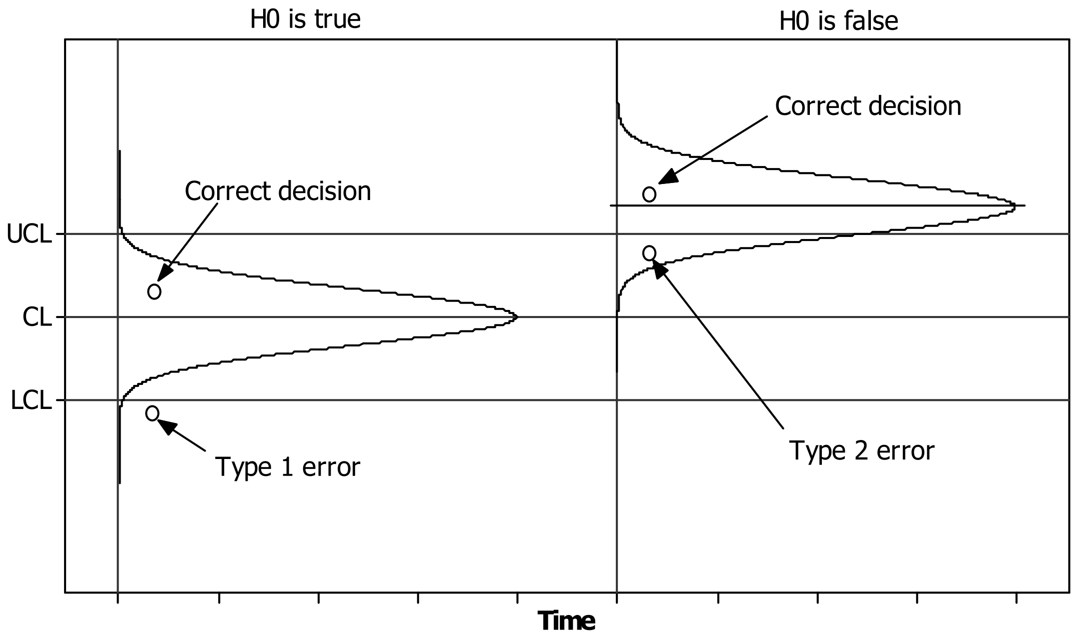
Decision Errors in Acceptance Sampling

The hypotheses are:

H_0*: the lot is good* versus H_A*: the lot is bad*



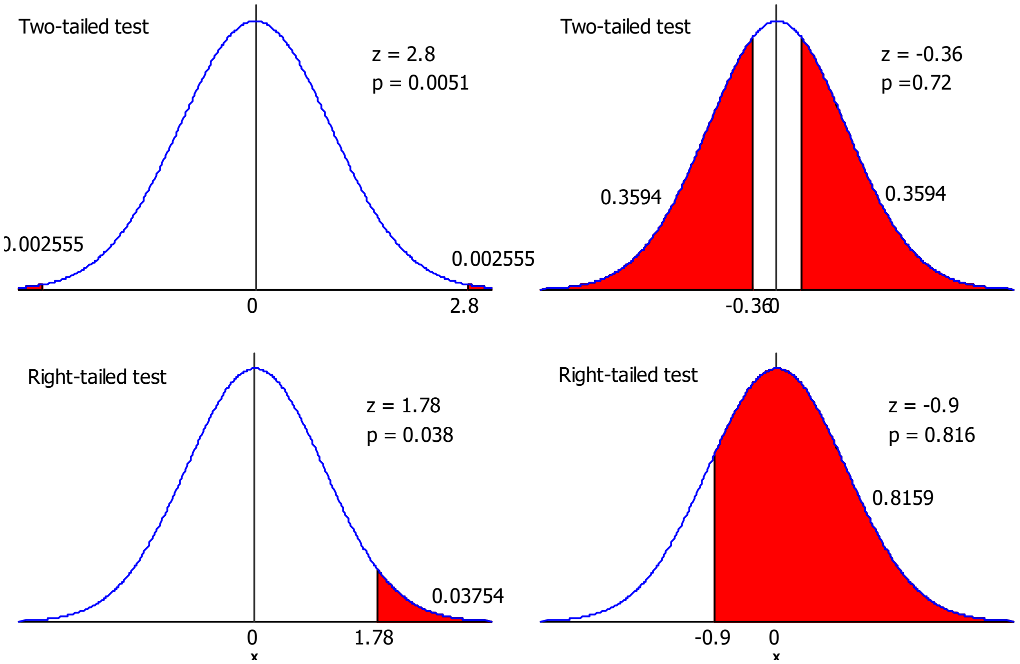
Decision Errors in SPC



Hypothesis Test p Values

- p values provide a concise and universal way of communicating statistical significance.
- The p value of a hypothesis test is the probability of obtaining the observed experimental result or something more extreme if the null hypothesis was true.
- p values are compared directly to α (typically $\alpha = 0.05$ or $\alpha = 0.01$) to make decisions about accepting or rejecting the null hypothesis.
  - If $p \geq \alpha$ accept H_0, that is, the data support the null hypothesis.
  - If $p < \alpha$ reject H_0, that is, the data don't support the null hypothesis.
- For two tailed hypothesis tests, the p value corresponds to the area in the two tails of the sampling distribution of the test statistic outside of the value obtained for the test statistic.
- For one tailed hypothesis tests, the p value corresponds to the area in one tail of the sampling distribution of the test statistic outside of the value obtained for the test statistic.

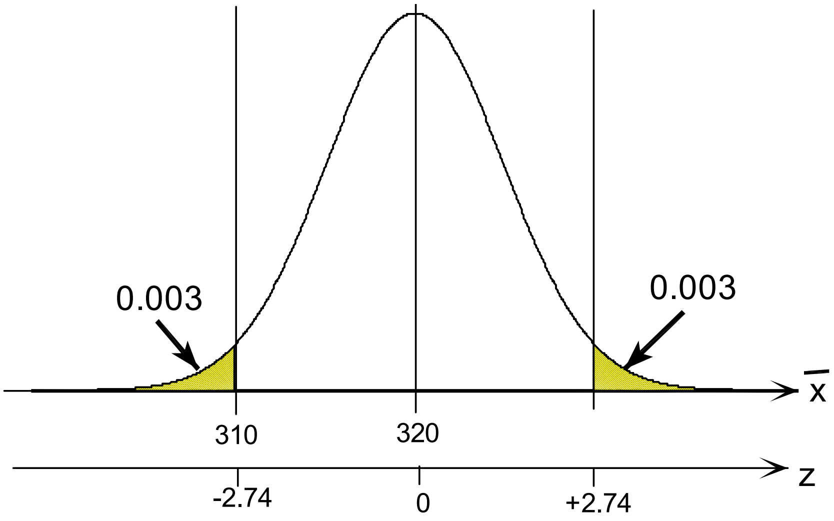
p Values



p Values

Example: Find the p value for Example A.

Solution: Since $z_{0.003} = 2.74$ the p value for this Example is $p = 2(0.003) = 0.006$. Because $(p = 0.006) < (\alpha = 0.05)$ we must reject the claim $H_0 : \mu = 320$.



Type 1 Error

Example: In a hypothesis test for $H_0 : \mu = 18$ vs. $H_A : \mu \neq 18$ the null hypothesis is accepted if the mean of a sample of size $n = 16$ falls within the interval $17.2 \leq \bar{x} \leq 18.8$. The population being sampled is normal and has $\sigma = 1.5$. Find the probability of committing a Type 1 error.

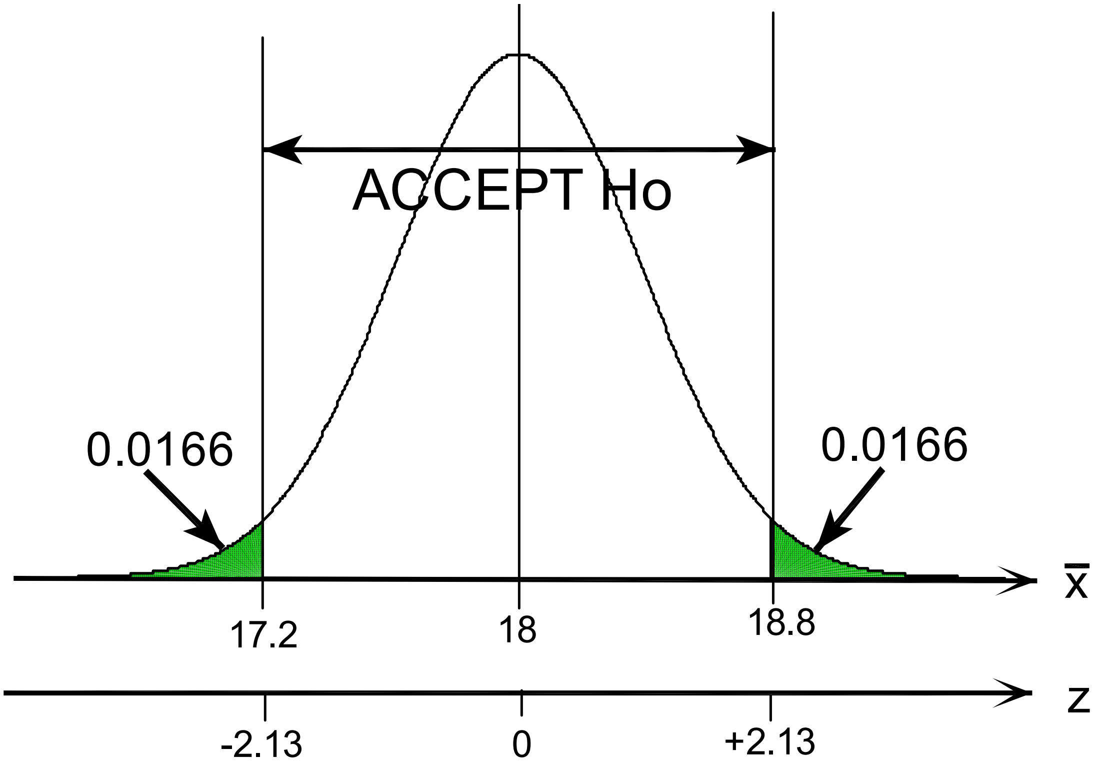
Solution: Type 1 errors occur when the null hypothesis is really true but a sample is obtained with a mean that falls outside of the acceptance interval. The probability of $\bar{x}$s falling inside the acceptance interval is:

$$\Phi(\mu - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu + z_{\alpha/2}\sigma_{\bar{x}}; \mu = \mu_0, \sigma_{\bar{x}}) = 1 - \alpha$$

where μ_0 is the hypothesized mean in the null hypothesis (i.e. $\mu_0 = 18$). If we check the upper decision limit $(UDL = 18.8)$ we have $\mu + z_{\alpha/2}\sigma_{\bar{x}} = UDL$ and solving for $z_{\alpha/2}$:

$$z_{\alpha/2} = \frac{UDL - \mu}{\sigma_{\bar{x}}} = \frac{18.8 - 18.0}{1.5/\sqrt{16}} = 2.13$$

Similarly, the lower decision limit $(LDL = 17.2)$ corresponds to $-z_{0.0166} = -2.13$. Since $z_{0.0166} = 2.13$ the probability of committing a Type 1 error is $\alpha = 2(0.0166) = 0.033$.



Type 2 Error

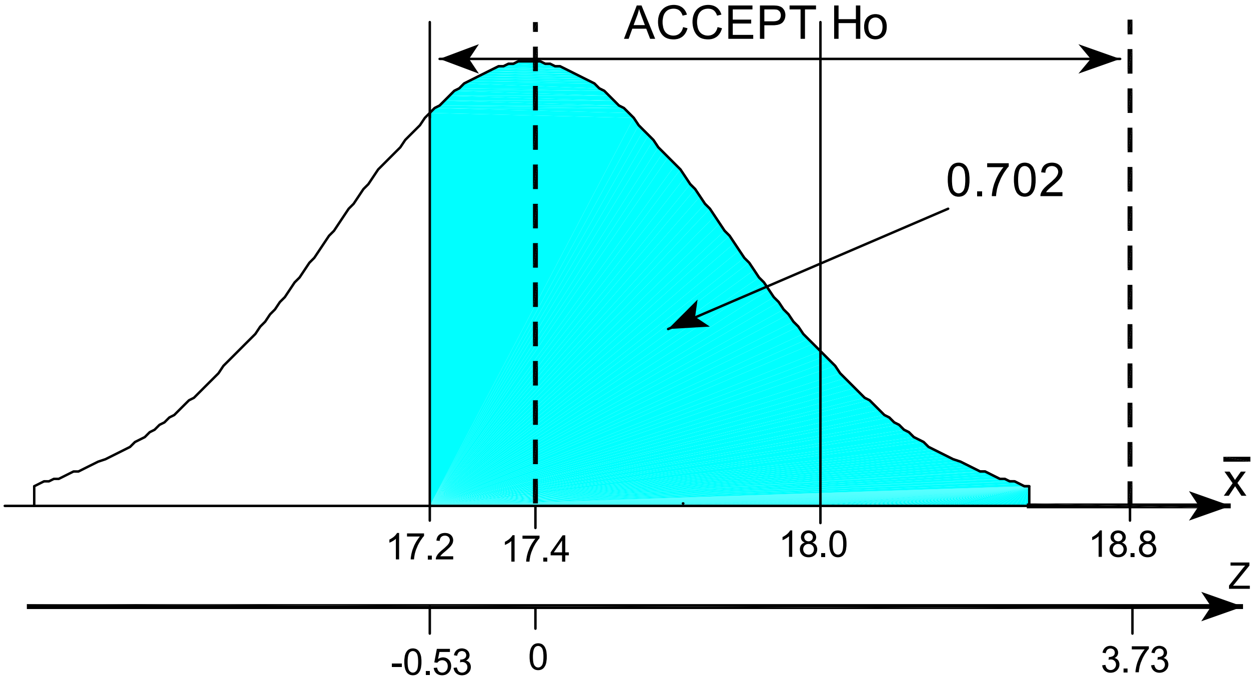
Example: In a hypothesis test for $H_0 : \mu = 18$ vs. $H_A : \mu \neq 18$ the null hypothesis is accepted if the mean of a sample of size $n = 16$ falls within the interval $17.2 \leq \bar{x} \leq 18.8$. The population being sampled is normal and has $\sigma = 1.5$. Find the probability of committing a Type 2 error when the true mean is $\mu = 17.4$.

Solution: Type 2 errors occur when the null hypothesis is really false but the test returns an erroneous *accept* H_0 result. The probability of committing a Type 2 error when the null hypothesis is really false is:

$$\beta = \Phi(\mu - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu + z_{\alpha/2}\sigma_{\bar{x}}; \mu \neq \mu_0; \sigma_{\bar{x}})$$

In this case we have:

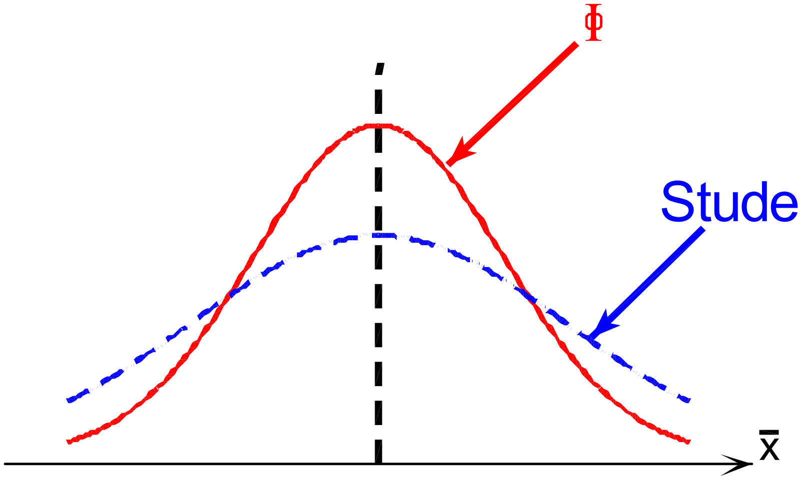
$$\begin{aligned}\beta &= \Phi(\mu - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu + z_{\alpha/2}\sigma_{\bar{x}}; \mu \neq \mu_0; \sigma_{\bar{x}}) \\ &= \Phi(17.2 < \bar{x} < 18.8; \mu = 17.4; 0.375) \\ &= \Phi(-0.53 < z < 3.73) \\ &= 1.00 - 0.298 \\ &= 0.702\end{aligned}$$



One Sample t Test

Example B: Test the hypothesis $H_0 : \mu = 440$ vs. $H_A : \mu \neq 440$ if a sample of size $n = 10$ yields $\bar{x} = 442$ and $s = 5.1$. Assume that the distribution of x is normal and work at a 5% significance level.

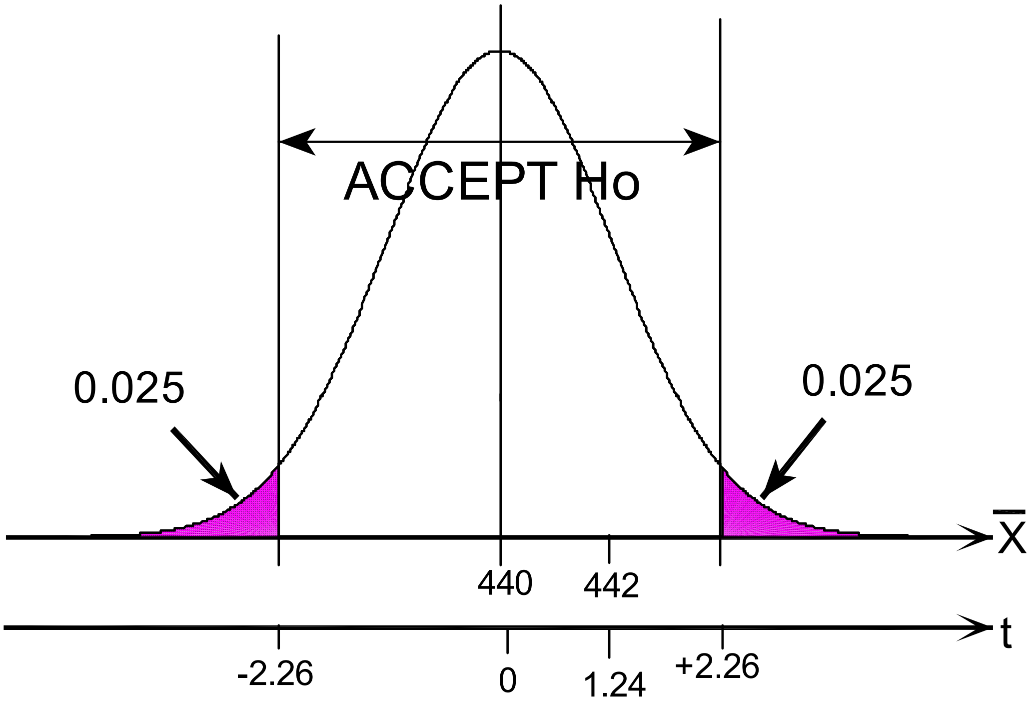
Solution: This is a hypothesis test for one sample mean but the central limit theorem doesn't apply because we don't know σ and don't have a good estimate for it. So ...



Solution: Since we don't know the true population standard deviation we must use Student's t distribution to characterize the distribution of sample means. From Student's t distribution with $n-1 = 9$ degrees of freedom we have $t_{0.025,9} = 2.26$ so the acceptance interval for H_0 is $-2.26 \leq t \leq 2.26$. The value of the t statistic is:

$$\begin{aligned} t &= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \\ &= \frac{442-440}{5.1/\sqrt{10}} = 1.24 \end{aligned}$$

Since the sample mean falls so close to the hypothesized mean and easily inside the acceptance interval we must accept the null hypothesis $H_0 : \mu = 440$.



Example: Find the p value for Example B.

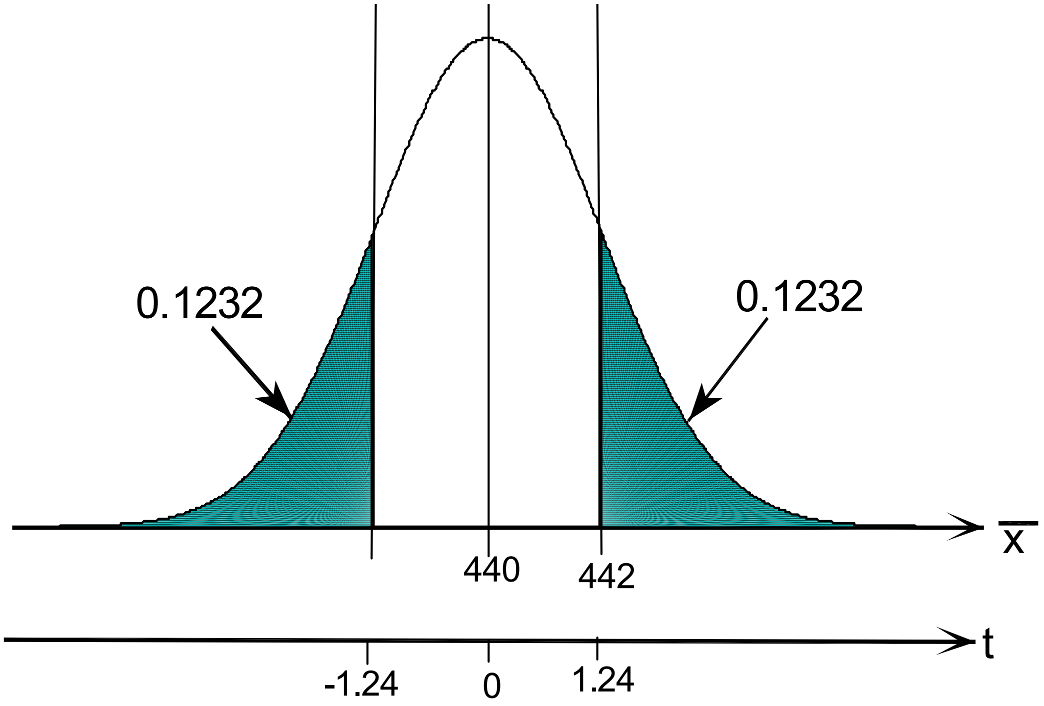
Solution: The p value is given by:

$$1 - p = P(-1.24 \leq t \leq 1.24)$$

where the Student's t distribution has $n - 1 = 9$ degrees of freedom. Generally it would be necessary to interpolate in a t table to estimate the true p value but MINITAB or Excel gives the exact p value:

$$p = 2(0.1232) = 0.246$$

Since $(p = 0.246) > (\alpha = 0.05)$ we must accept $H_0 : \mu = 440$.



Confidence Interval for μ When σ is Unknown

- σ unknown
- Distribution of x is normal
- The confidence interval for the population mean based on a sample of size n taken from a normal population which yields $\bar{x}$ and s is given by:

  $$P(\bar{x} - t_{\alpha/2}s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}s/\sqrt{n}) = 1 - \alpha$$

  where $t_{\alpha/2}$ comes from Student's t distribution with $\nu = n - 1$ degrees of freedom.

Confidence Interval

Example: Construct the 95% confidence interval for the true population mean for the situation in Example B.

Solution: The confidence interval for μ is:

$$P(\bar{x} - t_{\alpha/2}s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}s/\sqrt{n}) = 1 - \alpha$$

$$P\left(442 - 2.26 \times 5.1/\sqrt{10} < \mu < 442 + 2.26 \times 5.1/\sqrt{10}\right) = 0.95$$

$$P(438.4 < \mu < 445.6) = 0.95$$

That is, we can be 95% confident that the true population mean falls in the interval from 438.4 to 445.6.

This confidence interval demonstrates the relationship between confidence intervals and hypothesis tests: *a confidence interval for the mean is the set of population means for which the null hypothesis must be accepted*, so because the example's confidence interval contains $\mu = 440$ we know that we have to accept the null hypothesis $H_0 : \mu = 440$.

Two Independent Sample t Test

Data: Two samples of measurement data of size n_1 and n_2 from independent normal populations with equal variances $(\sigma_1^2 = \sigma_2^2)$.

Hypotheses Tested:

- $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 < \mu_2$
- $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$

Test Statistic:

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

where

$$s_{pooled} = \sqrt{\frac{\sum \epsilon_{1i}^2 + \sum \epsilon_{2i}^2}{n_1 - 1 + n_2 - 1}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Critical Values:

- For $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ accept H_0 iff $-t_{\alpha/2,n_1+n_2-2} < t < t_{\alpha/2,n_1+n_2-2}$
- For $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 < \mu_2$ accept H_0 iff $t > -t_{\alpha,n_1+n_2-2}$
- For $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$ accept H_0 iff $t < t_{\alpha,n_1+n_2-2}$

Behrens-Fisher Problem:

- Behrens and Fisher asked how to perform the two-sample t test when the two variances are not equal.
- The solution is called the Satterthwaite or Welch method.
- The Satterthwaite method is in excellent agreement with the assumed-equal-variances method when the variances are equal so we usually use the Satterthwaite method at all times.
- The Satterthwaite method is painful to calculate so it’s usually done with software.

Two Independent Sample t Test

Example: Samples are drawn from two processes to compare their means. The first sample yields $n_1 = 10$, $\bar{x}_1 = 278$, and $s_1 = 4.4$. The second sample yields $n_2 = 12$, $\bar{x}_2 = 280$, and $s_2 = 5.9$. Test the hypotheses $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level.

Solution: The test statistic for the two independent sample t test is:

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

where

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

For the given data:

$$s_{pooled} = \sqrt{\frac{(10 - 1)(4.4)^2 + (12 - 1)(5.9)^2}{10 + 12 - 2}} = 5.28$$

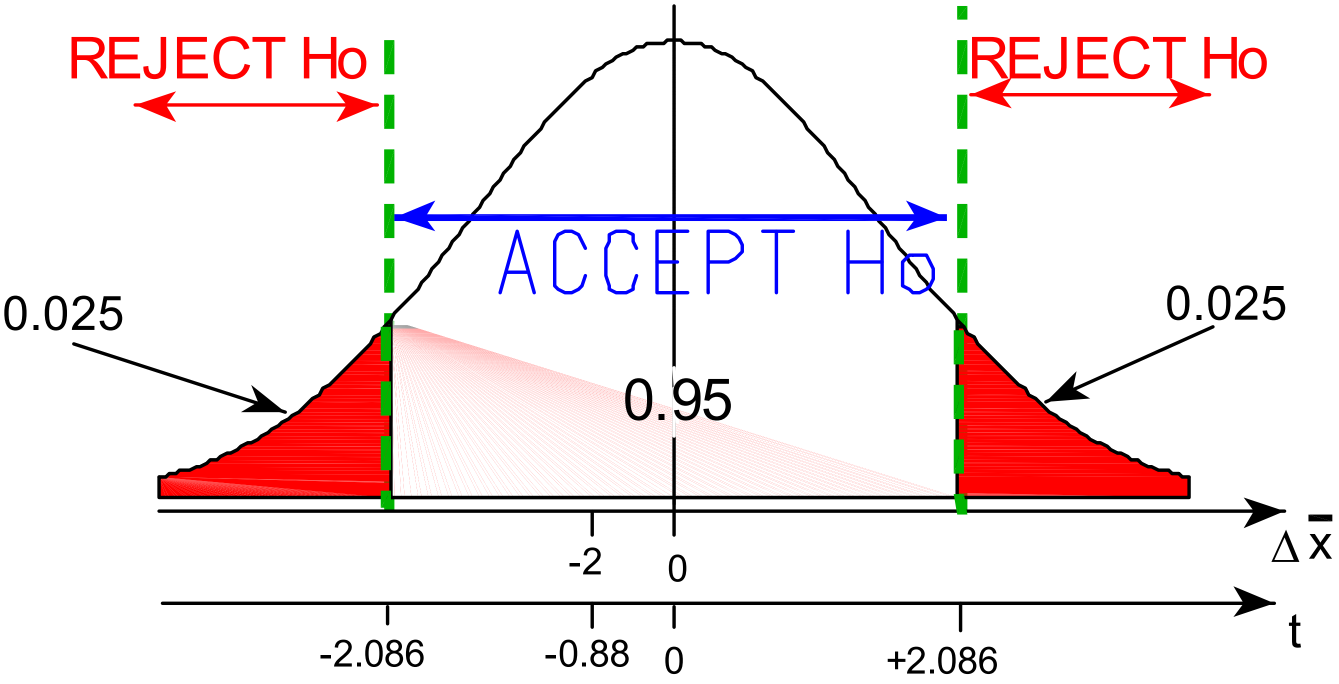
so the test statistic is:

$$t = \frac{278 - 280}{5.28\sqrt{\frac{1}{10} + \frac{1}{12}}} = -0.88$$

Since $t_{\alpha/2, n_1+n_2-2} = t_{0.025,20} = 2.086$ the acceptance interval for the null hypothesis is:

Accept H_0 iff $-2.086 \le t \le 2.086$

The test statistic $t = -0.88$ falls within this interval so we must accept the null hypothesis and conclude that $\mu_1 = \mu_2$.



Confidence Interval for the Difference Between Two Population Means

- σ_1 and σ_2 are equal but unknown
- Both samples come from normal populations
- The confidence interval for the difference between two population means is given by:

$$P(\Delta\bar{x} - t_{\alpha/2} s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \Delta\mu < \Delta\bar{x} + t_{\alpha/2} s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}) = 1 - \alpha$$

where

$$\Delta\bar{x} = \bar{x}_1 - \bar{x}_2$$

$$\Delta\mu = \mu_1 - \mu_2$$

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

and $t_{\alpha/2}$ has $\nu = n_1 + n_2 - 2$ degrees of freedom.

Confidence Interval

Example: Two samples yield the following values:

$n_1 = 8, \bar{x}_1 = 18.8, s_1 = 1.5$ and $n_2 = 10, \bar{x}_2 = 15.6, s_2 = 2.4$

Construct the 95% confidence interval for the difference between the population means.

Solution: We must assume that the populations being sampled are normal and that the variances are equal. The pooled standard deviation is:

$$s_{pooled} = \sqrt{\frac{(8-1)(1.5)^2 + (10-1)(2.4)^2}{8 + 10 - 2}} = 2.06$$

With $\Delta\bar{x} = 18.8 - 15.6 = 3.2$ and with $\nu = 8 + 10 - 2 = 16$ degrees of freedom we have $t_{0.025,16} = 2.12$. The confidence interval is:

$$P(3.2 - 2.12 \times 2.06 \times \sqrt{\frac{1}{8} + \frac{1}{10}} < \Delta\mu < 3.2 + 2.12 \times 2.06 \times \sqrt{\frac{1}{8} + \frac{1}{10}}) = 0.95$$

$$P(1.13 < \Delta\mu < 5.27) = 0.95$$

Paired Sample t Test

Data: n paired samples (x_{1i}, x_{2i}) of measurement data taken from normal populations. The data pairs are ”before and after” type.

Test Statistic: The quantities of interest are the signed differences between the paired observations:

$$\Delta x_i = x_{1i} - x_{2i}$$

The mean and standard deviation of these differences are required:

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

and

$$s = \sqrt{\frac{\sum (\Delta x_i - \overline{\Delta x})^2}{n-1}}$$

The test statistic is:

$$t = \frac{\overline{\Delta x}}{s/\sqrt{n}}$$

Hypotheses Tested:

- $H_0 : \mu = 0$ vs. $H_A : \mu \neq 0$
- $H_0 : \mu = 0$ vs. $H_A : \mu < 0$
- $H_0 : \mu = 0$ vs. $H_A : \mu > 0$

Critical Values:

- For $H_0 : \mu = 0$ vs. $H_A : \mu \neq 0$ accept H_0 iff $-t_{\alpha/2,n-1} < t < t_{\alpha/2,n-1}$
- For $H_0 : \mu = 0$ vs. $H_A : \mu < 0$ accept H_0 iff $t > -t_{\alpha,n-1}$
- For $H_0 : \mu = 0$ vs. $H_A : \mu > 0$ accept H_0 iff $t < t_{\alpha,n-1}$

Paired Sample *t* Test

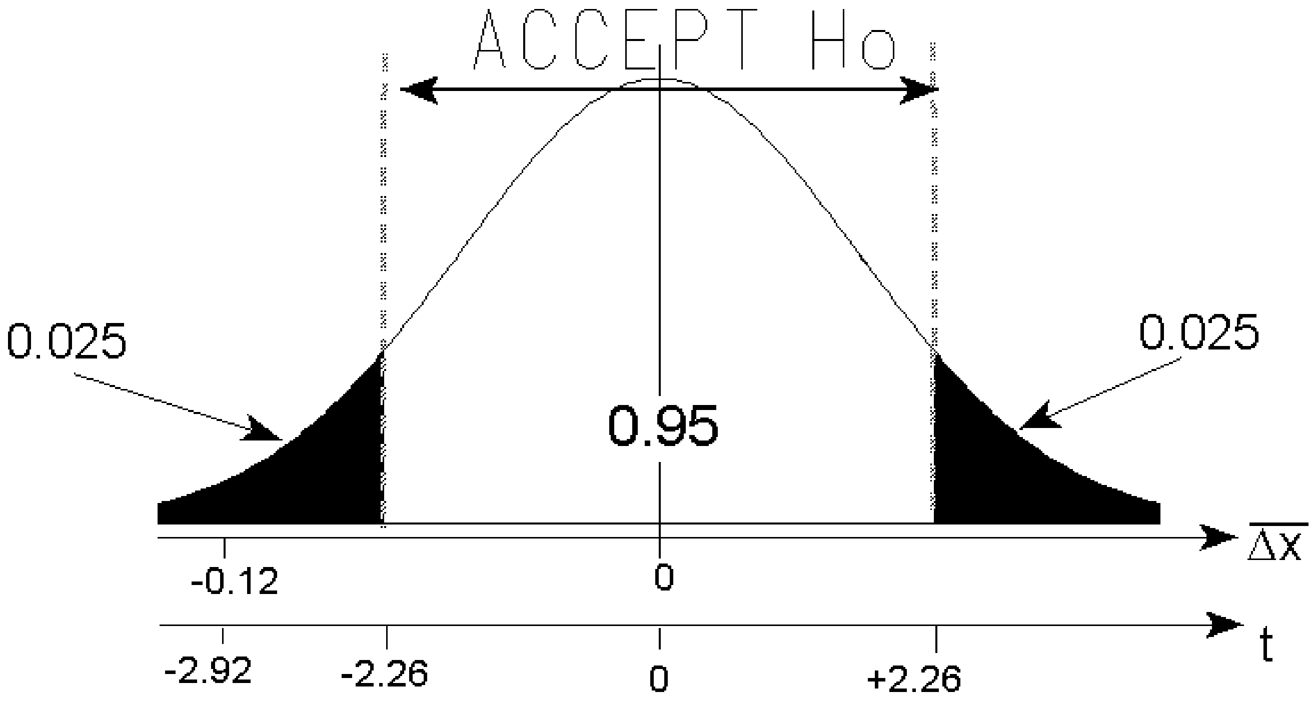
Example: The following table shows measurements taken by two operators on the same 10 parts. Determine if there is evidence that they are getting different readings at the 5% significance level.

Part Number	1	2	3	4	5	6	7	8	9	10
Operator 1	2.4	2.8	3.1	2.7	3.0	2.5	2.2	4.3	3.8	3.4
Operator 2	2.6	2.9	3.4	2.7	2.9	2.7	2.3	4.4	4.1	3.4

Solution: The differences between the paired readings are shown below:

Part Number	1	2	3	4	5	6	7	8	9	10
Operator 1	2.4	2.8	3.1	2.7	3.0	2.5	2.2	4.3	3.8	3.4
Operator 2	2.6	2.9	3.4	2.7	2.9	2.7	2.3	4.4	4.1	3.4
Δx_i	-0.2	-0.1	-0.3	0.0	0.1	-0.2	-0.1	-0.1	-0.3	0.0

The mean of the differences is $\overline{\Delta x} = -1.2/10 = -0.12$ and the standard deviation of the differences is $s = 0.13$. The test statistic is $t = \frac{-0.12}{0.13/\sqrt{10}} = -2.92$. If the hypotheses tested are $H_0 : \mu = 0$ vs. $H_A : \mu \neq 0$ then the critical value of the test statistic is $t_{0.025,9} = 2.26$ and the acceptance interval for the null hypothesis is $-2.26 < t < 2.26$. Since $t = -2.92$ falls outside this interval we must reject H_0 and conclude that there is a statistically significant difference between the two operators.



Distribution of Sample Variances

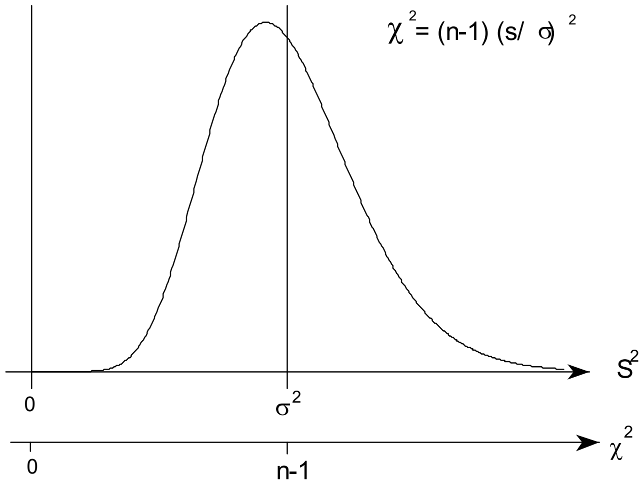
If repeated samples of size n are drawn from a normal population and the sample variances are determined, then the distribution of sample variances is chi-square with $n-1$ degrees of freedom.

Notes About the χ^2 Distribution

- Always skewed right
- Measurement units are transformed to standard units by

$$\chi^2 = (n-1)\left(\frac{s}{\sigma}\right)^2$$

- Mean is $\mu_{\chi^2} = n-1$
- Changes shape as n changes
- Becomes normal (Φ) as $n \to \infty$
- Used to construct confidence intervals for the population variance
- Used to determine accept/reject limits for hypothesis tests based on one sample variance
- Variances are very very noisy



Confidence Interval for σ^2

The two sided confidence interval for σ^2 determined from the sample variance s^2 with a sample of size n is given by:

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

where the chi-square distribution has $n-1$ degrees of freedom.

(Note: The subscript on χ^2 indicates the left tail area under the χ^2 distribution. Some texts index χ^2 tables by the right tail area instead.)

Confidence Interval for σ^2

Example: A random sample of size $n = 18$ taken from a normal population yields a standard deviation of $s = 5.4$. Determine a 95% confidence interval for the population standard deviation.

Solution: The confidence interval is given by:

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

From the χ^2 tables we find $\chi^2_{0.025,17} = 7.56$ and $\chi^2_{0.975,17} = 30.19$. The required confidence interval for the population variance is:

$$P\left(\frac{17}{30.19}(5.4)^2 < \sigma^2 < \frac{17}{7.56}(5.4)^2\right) = 0.95$$

$$P(16.4 < \sigma^2 < 65.6) = 0.95$$

$$P(4.05 < \sigma < 8.10) = 0.95$$

Hypothesis Test for One Variance

The hypotheses to be tested are $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_A : \sigma^2 \neq \sigma_0^2$. The distribution of sample variances suggests the following form for the acceptance interval for H_0:

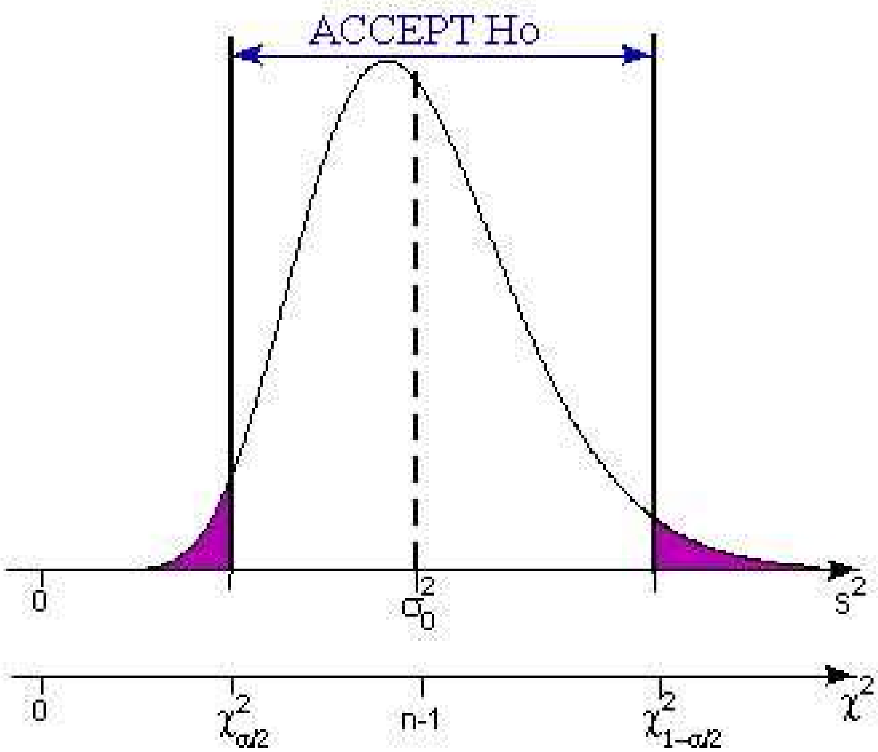
$$P\left(\frac{\chi^2_{\alpha/2}}{n-1}\sigma_0^2 < s^2 < \frac{\chi^2_{1-\alpha/2}}{n-1}\sigma_0^2\right) = 1 - \alpha$$

However, it is generally easier to make the decision on the basis of the test statistic:

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

with acceptance interval for the null hypothesis given by:

$$P(\chi^2_{\alpha/2} < \chi^2 < \chi^2_{1-\alpha/2}) = 1 - \alpha$$



Hypothesis Test for One Variance

Example: A random sample of size $n = 25$ taken from a normal population yields $s^2 = 75$. Test the hypotheses $H_0 : \sigma^2 = 50$ vs. $H_A : \sigma^2 \neq 50$ at the $\alpha = 0.05$ significance level.

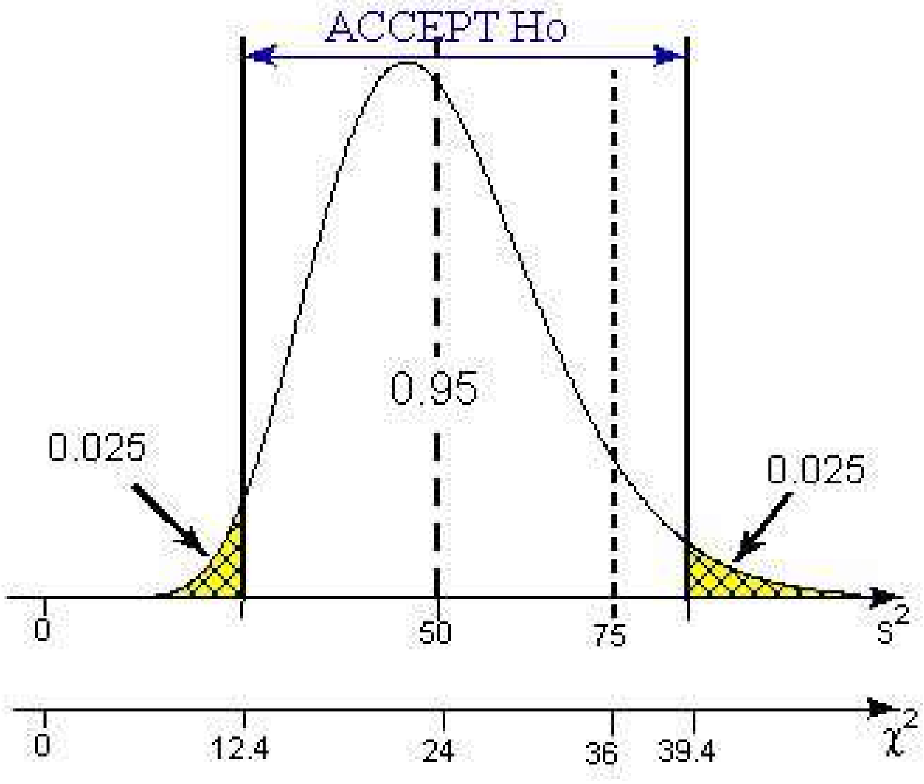
Solution: The χ^2 statistic is:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(24)75}{50} = 36$$

From the χ^2 table we have $\chi^2_{0.025,24} = 12.4$ and $\chi^2_{0.975,24} = 39.4$ so the acceptance interval for H_0 is:

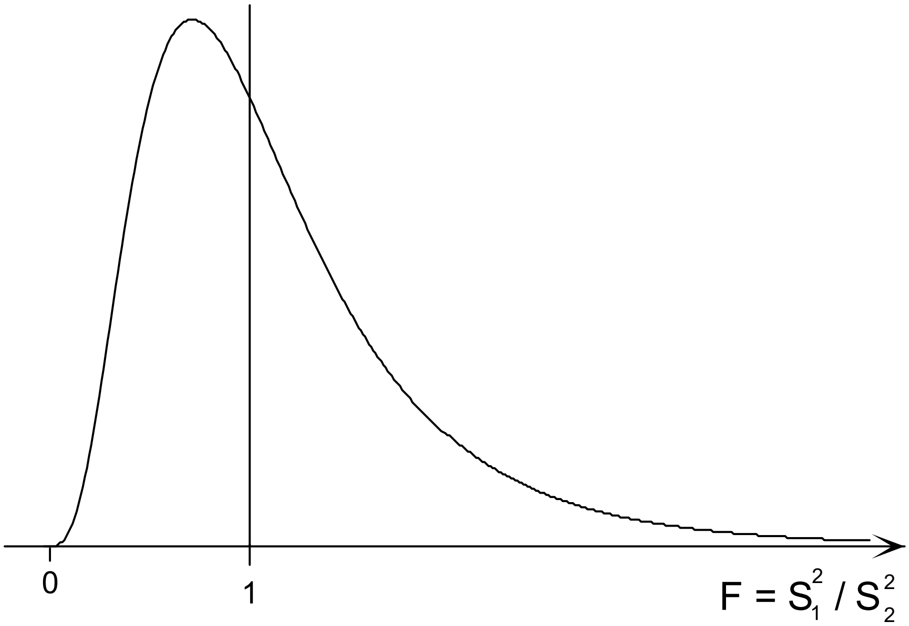
$$\begin{aligned} P(\chi^2_{0.025} < \chi^2 < \chi^2_{0.975}) &= 0.95 \\ P(12.4 < \chi^2 < 39.4) &= 0.95 \end{aligned}$$

Since $\chi^2 = 36$ falls easily inside of the acceptance interval we must accept $H_0 : \sigma^2 = 50$.



Distribution of the Ratio of Two Sample Variances

If two samples of size n_1 and n_2 are drawn from normal populations that have equal population variances, then the ratio of their sample variances $F = s_1^2/s_2^2$ follows the F distribution with $n_1 - 1$ and $n_2 - 1$ numerator and denominator degrees of freedom, respectively.



Notes About the F Distribution

- Always skewed right
- Mean is $\mu_F = 1$
- Changes shape as n_1 and n_2 change
- Used to determine accept/reject limits for hypothesis tests comparing two sample variances
- $F = s_1^2/s_2^2$ is usually constructed such that $s_1 > s_2$ and only right tail F values are indexed in the tables, sometimes by right and sometimes by left tail area
- Variances are very very noisy

Hypothesis Test for Two Variances

Example: Random samples of size $n_1 = 12$ and $n_2 = 16$ are drawn from two populations. The sample standard deviations are found to be $s_1 = 145$ and $s_2 = 82$. Test to see if there is evidence that the population variances are equal at the $\alpha = 0.05$ significance level.

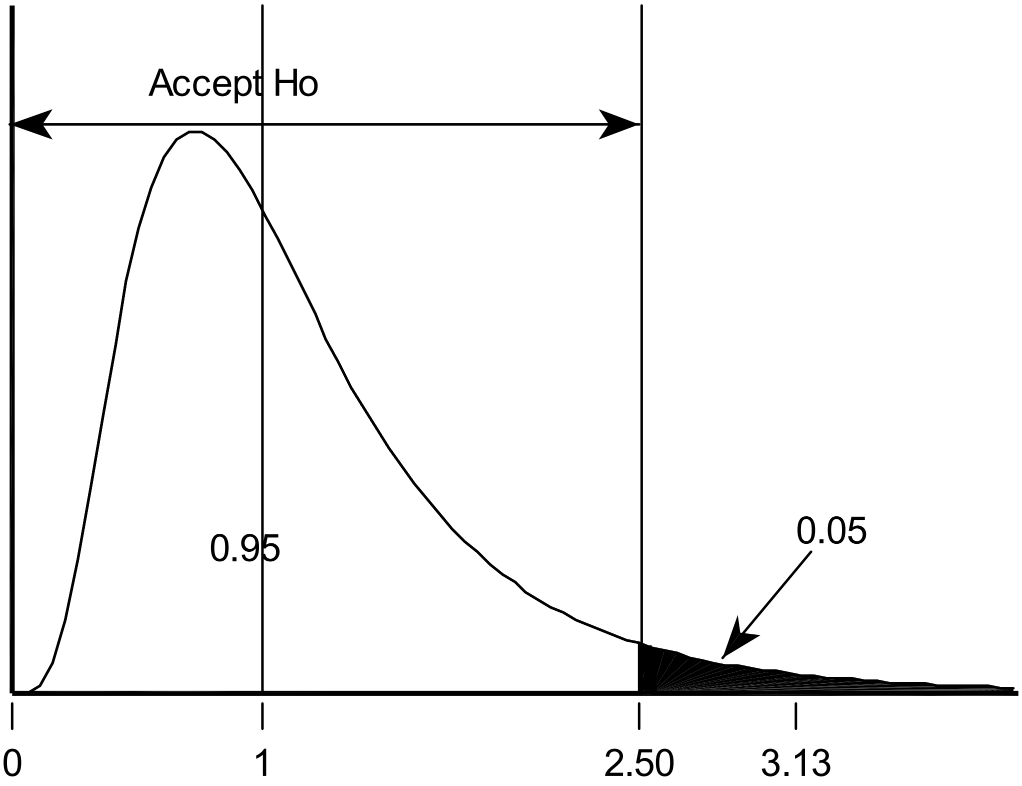
Solution: The hypotheses to be tested are $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_A : \sigma_1^2 > \sigma_2^2$. The acceptance interval for the null hypothesis is given by:

$$P\left(0 < \frac{s_1^2}{s_2^2} < F_{1-\alpha}\right) = 1 - \alpha$$

From the F tables with 11 numerator and 15 denominator degrees of freedom we find $F_{0.95} = 2.51$. The F statistic is given by:

$$\begin{aligned} F &= s_1^2/s_2^2 \\ &= (145/82)^2 \\ &= 3.13 \end{aligned}$$

Since $F = 3.13$ falls outside the acceptance interval we must reject H_0 and conclude that there is evidence that the two populations being sampled have different variances.



Summary of Hypothesis Testing Methods

	One	**Paired**	**Two**	**Many**
Mean	z or t sign test Wilcoxon SRT	$t\ (\Delta x_i = x_{1i} - x_{2i})$ paired sample sign Wilcoxon paired SR	$t\ (\sigma_1 = \sigma_2)$ $t\ (\sigma_1 \neq \sigma_2)$ Tukey's quick test Boxplot slippage Mann-Whitney	ANOVA MCT(e.g., Tukey, …) Kruskal-Wallis Mood's median Regression
Standard Deviation	χ^2		F Levene squared ranks	Bartlett Hartley's $F_{\max}$ Levene ANOVA or regr. of $\log(s^2)$
Proportion	exact binomial Larson's nomogram normal approx.	McNemar	Fisher's exact normal approx.	χ^2 ANOVA of $\sin^{-1}\left(\sqrt{p_i}\right)$ Cochran's Q Binary logistic regression
Count	exact Poisson normal approx.		exact binomial normal approx. F	ANOVA of $\sqrt{x_i}$ log-linear models Poisson regression
Dist. Shape	probability plot χ^2goodness of fit Shapiro-Wilk Anderson-Darling Kolmogorov		Smirnov	

Summary of Sampling Distributions and Confidence Intervals

Quantity	Condition	Sampling Distribution
Mean	CLT^2	$\Phi(\mu - z_{\alpha/2}\sigma_{\bar{x}} < \bar{x} < \mu + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$
Mean	σ unknown, $\Phi(x)^1$	$P(\mu - t_{\alpha/2}s/\sqrt{n} < \bar{x} < \mu + t_{\alpha/2}s/\sqrt{n}) = 1 - \alpha$
Variance	$\Phi(x)$	$P\left(\frac{\chi^2_{\alpha/2}}{n-1}\sigma^2 < s^2 < \frac{\chi^2_{1-\alpha/2}}{n-1}\sigma^2\right) = 1 - \alpha$
Standard Deviation	$\Phi(x)$, $n > 30$	$P\left[\left(1 - \frac{z_{\alpha/2}}{\sqrt{2n}}\right)\sigma < s < \left(1 + \frac{z_{\alpha/2}}{\sqrt{2n}}\right)\sigma\right] = 1 - \alpha$
Ratio of Variances	$\Phi(x_1)$, $\Phi(x_2)$	$P\left(F_{1-\alpha/2} < \frac{s_1^2}{s_2^2} < F_{\alpha/2}\right) = 1 - \alpha$
Proportion	n large	$P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$
Proportion	n large	NA

Quantity	Condition	Confidence Interval
Mean	CLT^2	$\Phi(\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} < \mu < \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$
Mean	σ unknown, $\Phi(x)^1$	$P(\bar{x} - t_{\alpha/2}s/\sqrt{n} < \mu < \bar{x} + t_{\alpha/2}s/\sqrt{n}) = 1 - \alpha$
Variance	$\Phi(x)$	$P\left(\frac{n-1}{\chi^2_{1-\alpha/2}}s^2 < \sigma^2 < \frac{n-1}{\chi^2_{\alpha/2}}s^2\right) = 1 - \alpha$
Standard Deviation	$\Phi(x)$, $n > 30$	$P\left[s/\left(1 + \frac{z_{\alpha/2}}{\sqrt{2n}}\right) < \sigma < s/\left(1 - \frac{z_{\alpha/2}}{\sqrt{2n}}\right)\right] = 1 - \alpha$
Ratio of Variances	$\Phi(x_1)$, $\Phi(x_2)$	NA
Proportion	n large	$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$
Proportion	n large	$P\left(0 < p < \frac{1}{2n}\chi^2_{1-\alpha,2(x+1)}\right) = 1 - \alpha$ where x is #failures

Notes:

1) $\Phi(x)$ means that the distribution of x is normal.

2) CLT (Central Limit Theorem) requires that $n \geq 30$ or $\Phi(x)$ with σ known. If σ is unknown or distribution of x is not normal then use $n \geq 30$ and $\sigma_x \simeq s$.

3) The χ^2 distribution is indexed by its left tail area. For example: $\chi^2_{0.05,10} = 3.94$ and $\chi^2_{0.95,10} = 18.3$.

4) The F distribution is indexed by its right tail area.

Test	H_0 **vs.** H_A : (H_0 **Acceptance Interval**)	**Test Statistic**
One Mean σ known	$\mu = \mu_0$ vs. $\mu \neq \mu_0$: $(-z_{\alpha/2} \leq z \leq z_{\alpha/2})$ $\mu = \mu_0$ vs. $\mu < \mu_0$: $(-z_{\alpha} \leq z < \infty)$ $\mu = \mu_0$ vs. $\mu > \mu_0$: $(-\infty < z \leq z_{\alpha})$	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
One Mean σ unknown	$\mu = \mu_0$ vs. $\mu \neq \mu_0$: $(-t_{\alpha/2} \leq t \leq t_{\alpha/2})$ $\mu = \mu_0$ vs. $\mu < \mu_0$: $(-t_{\alpha} \leq t < \infty)$ $\mu = \mu_0$ vs. $\mu > \mu_0$: $(-\infty < t \leq t_{\alpha})$	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ $\nu = n-1$
Two Means Independent Samples σs known	$\mu_1 = \mu_2$ vs. $\mu_1 \neq \mu_2$: $(-z_{\alpha/2} \leq z \leq z_{\alpha/2})$ $\mu_1 = \mu_2$ vs. $\mu_1 < \mu_2$: $(-z_{\alpha} \leq z < \infty)$ $\mu_1 = \mu_2$ vs. $\mu_1 > \mu_2$: $(-\infty < z \leq z_{\alpha})$	$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
Two Means Independent Samples σs unknown but equal	$\mu_1 = \mu_2$ vs. $\mu_1 \neq \mu_2$: $(-t_{\alpha/2} \leq t \leq t_{\alpha/2})$ $\mu_1 = \mu_2$ vs. $\mu_1 < \mu_2$: $(-t_{\alpha} \leq t < \infty)$ $\mu_1 = \mu_2$ vs. $\mu_1 > \mu_2$: $(-\infty < t \leq t_{\alpha})$	$t = \frac{\bar{x}_1-\bar{x}_2}{s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$ $\nu = n_1+n_2-2$
Two Means Independent Samples σs unknown, unequal	$\mu_1 = \mu_2$ vs. $\mu_1 \neq \mu_2$: $(-t_{\alpha/2} \leq t \leq t_{\alpha/2})$ $\mu_1 = \mu_2$ vs. $\mu_1 < \mu_2$: $(-t_{\alpha} \leq t < \infty)$ $\mu_1 = \mu_2$ vs. $\mu_1 > \mu_2$: $(-\infty < t \leq t_{\alpha})$	$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $\nu = \min(n_1-1, n_2-1)$ $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$
One Mean Paired Samples σ unknown	$\Delta\mu = 0$ vs. $\Delta\mu \neq 0$: $(-t_{\alpha/2} \leq t \leq t_{\alpha/2})$ $\Delta\mu = 0$ vs. $\Delta\mu < 0$: $(-t_{\alpha} \leq t < \infty)$ $\Delta\mu = 0$ vs. $\Delta\mu > 0$: $(-\infty < t \leq t_{\alpha})$	$\Delta x_i = x_{1i}-x_{2i}$ $t = \frac{\overline{\Delta x}}{s_{\Delta x}/\sqrt{n}}$ $\nu = n-1$
One Variance	$\sigma^2 = \sigma_0^2$ vs. $\sigma^2 \neq \sigma_0^2$: $(\chi^2_{\alpha/2} \leq \chi^2 \leq \chi^2_{1-\alpha/2})$ $\sigma^2 = \sigma_0^2$ vs. $\sigma^2 < \sigma_0^2$: $(0 < \chi^2 \leq \chi^2_{1-\alpha})$ $\sigma^2 = \sigma_0^2$ vs. $\sigma^2 > \sigma_0^2$: $(\chi^2_{\alpha} \leq \chi^2 < \infty)$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ $\nu = n-1$
Two Variances	$\sigma_1^2 = \sigma_2^2$ vs. $\sigma_1^2 \neq \sigma_2^2$: $(F_{1-\alpha/2} \leq F \leq F_{\alpha/2})$ $\sigma_1^2 = \sigma_2^2$ vs. $\sigma_1^2 < \sigma_2^2$: $(0 < F \leq F_{\alpha})$	$F = \frac{s_2^2}{s_1^2}$ $\nu_2 = n_2-1$ $\nu_1 = n_1-1$

Notes:

1) All populations being sampled are normally distributed.

2) The χ^2 distribution is indexed by left tail area.

3) The F distribution is indexed by right tail area.

Sample Size Calculations

- All data require some type of analysis
- Point estimates (e.g. $\bar{x}$ and s) are insufficient
- Appropriate analysis methods take into account estimation precision
- Appropriate analysis methods are:
  - Confidence intervals
  - Hypothesis tests
- After the method of analysis has been identified a sample size calculation can be done to determine the unique number of observations required to obtain practically significant results.
  - If the sample size is too small there may be excessive risks of type1 and type 2 errors.
  - If the sample size is too large the experiment will be oversensitive and wasteful of resources.

Confidence Interval for the Mean (σ known)

Conditions:

- σ known
- Distribution of x is Φ

Confidence Interval: The confidence interval will have the form:

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

where

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

The value of δ should be chosen so that a single management action is indicated over the range of the confidence interval.

Sample Size: To be $(1-\alpha)100\%$ confident that the population mean μ is within $\pm\delta$ of the sample mean $\bar{x}$, the required sample size is:

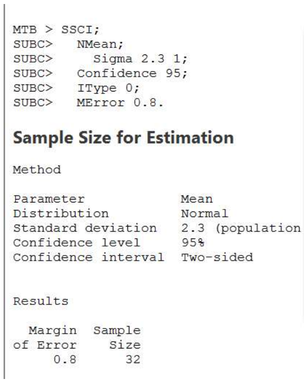
$$n = \left(\frac{z_{\alpha/2}\sigma}{\delta}\right)^2$$

Example: Find the sample size required to estimate the population mean to within ± 0.8 with 95% confidence if measurements are normally distributed with standard deviation $\sigma = 2.3$.

Solution: The sample size required is:

$$\begin{aligned} n &= \left(\tfrac{z_{0.025}\sigma}{\delta}\right)^2 \\ &= \left(\tfrac{1.96 \times 2.3}{0.8}\right)^2 \\ &= 31.8 \rightarrow 32 \end{aligned}$$

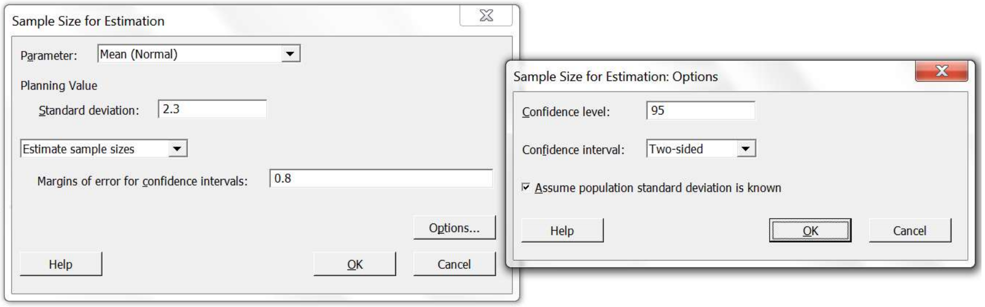
Or using MINITAB **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Mean** (**Normal**):



```
MTB > SSCI;
SUBC>   NMean;
SUBC>     Sigma 2.3 1;
SUBC>   Confidence 95;
SUBC>   IType 0;
SUBC>   MError 0.8.
```

Sample Size for Estimation

```
Method

Parameter            Mean
Distribution         Normal
Standard deviation   2.3 (population
Confidence level     95%
Confidence interval  Two-sided


Results

  Margin  Sample
of Error    Size
     0.8      32
```



Sample Size for Estimation

Parameter: Mean (Normal)

Planning Value

Standard deviation: 2.3

Estimate sample sizes

Margins of error for confidence intervals: 0.8

Sample Size for Estimation: Options

Confidence level: 95

Confidence interval: Two-sided

☑ Assume population standard deviation is known

Confidence Interval for the Mean (σ unknown)

- When σ is unknown it will be necessary to estimate it from the sample standard deviation and the t distribution will be used instead of the z distribution to calculate the confidence interval.
- But $t_{\alpha/2}$ depends on the sample size so our sample size equation for n is transcendental, i.e. has inseparable n dependencies on both sides of the equation so the sample size must be found by iterating.

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:

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

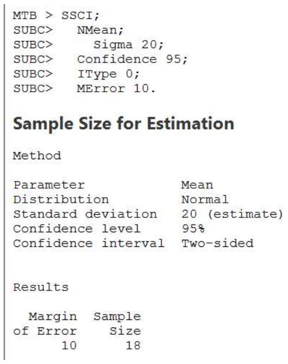
With $n = 16$, $\nu = 15$, and $t_{0.025} = 2.13$ so

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

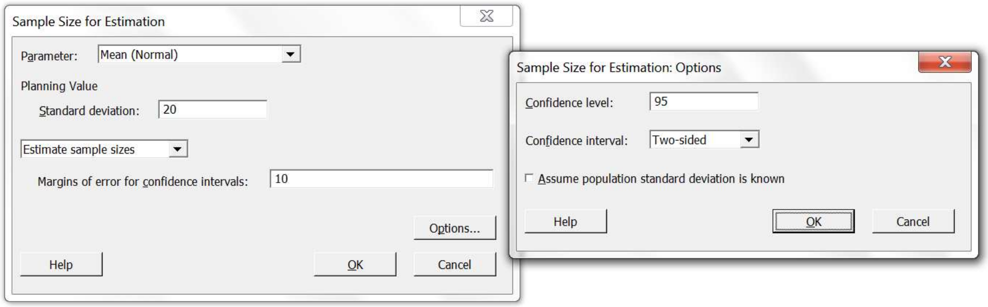
Eventually, with $n = 18$, $\nu = 17$, and $t_{0.025} = 2.11$:

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

Or using MINITAB **Stat**> **Power and Sample Size**> **Sample Size for Estimation**> **Mean** (**Normal**):



```
MTB > SSCI;
SUBC>   NMean;
SUBC>     Sigma 20;
SUBC>   Confidence 95;
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
```



Sample Size for Estimation

Parameter: Mean (Normal)

Planning Value

Standard deviation: 20

Estimate sample sizes

Margins of error for confidence intervals: 10

Sample Size for Estimation: Options

Confidence level: 95

Confidence interval: Two-sided

☐ Assume population standard deviation is known

Confidence Interval for the Difference Between Two Population Means

Conditions:

- σ_1 and σ_2 are known and equal
- Distributions of x_1 and x_2 are Φ

Confidence Interval: The confidence interval for $\Delta\mu = \mu_1 - \mu_2$ is

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

where

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

and $\Delta\bar{x} = \bar{x}_1 - \bar{x}_2$.

Sample Size: To be $(1-\alpha)100\%$ confident that the difference between two population means is within $\pm\delta$ of the difference in the sample means, the required sample size is:

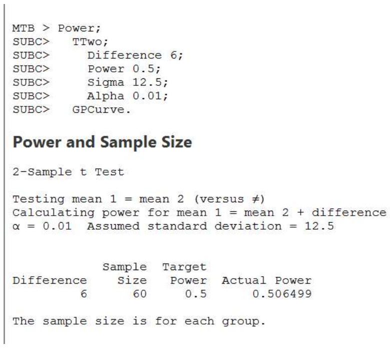
$$n = 2\left(\frac{z_{\alpha/2}\sigma}{\delta}\right)^2$$

Example: What sample size should be used to determine the difference between two population means to within ± 6 of the estimated difference to 99% confidence. The populations are normal and both have standard deviation $\sigma = 12.5$.

Solution: The required sample size is:

$$\begin{aligned} n &= 2\left(\frac{z_{\alpha/2}\sigma}{\delta}\right)^2 \\ &= 2\left(\frac{2.575\times 12.5}{6}\right)^2 \\ &= 57.6 \to 58 \end{aligned}$$

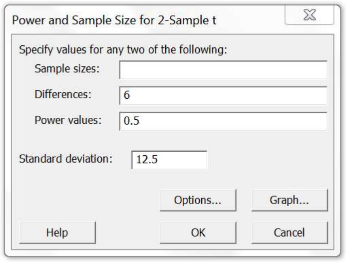
MINITAB does not offer a sample size calculation for the confidence interval for the difference between two population means but the **Stat**> **Power and Sample Size**> **2-Sample t** menu can be tricked into doing the calculation.



```
MTB > Power;
SUBC>   TTwo;
SUBC>     Difference 6;
SUBC>     Power 0.5;
SUBC>     Sigma 12.5;
SUBC>     Alpha 0.01;
SUBC>   GPCurve.
```

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus ≠)
Calculating power for mean 1 = mean 2 + difference
α = 0.01  Assumed standard deviation = 12.5


            Sample  Target
Difference    Size   Power  Actual Power
         6      60     0.5      0.506499

The sample size is for each group.
```



Power and Sample Size for 2-Sample t

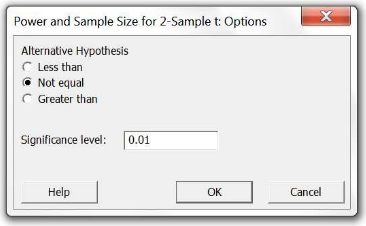
Specify values for any two of the following:

Sample sizes:

Differences: 6

Power values: 0.5

Standard deviation: 12.5



Power and Sample Size for 2-Sample t: Options

Alternative Hypothesis

○ Less than

● Not equal

○ Greater than

Significance level: 0.01

Input Information for the Sample Size Calculation

- 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

Issues in 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. For example, halving the value of the confidence interval half-width will quadruple the sample size.
- 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.

Sample Size Calculations for Hypothesis Tests

- When determining sample size for hypothesis tests it is necessary to specify the conditions and probabilities associated with Type 1 and Type 2 errors.
- The *power* of a test given by:

  $$\Pi = 1 - \beta$$

  is the probability of rejecting H_0 when H_A is true.
- A value of power is always associated with a corresponding value of *effect size* δ - the smallest practically significant difference between the population parameter under H_0 and H_A that the experiment should detect with probability Π.
- In all sample size calculations round n up to the nearest integer value.

Sample Size for a One-Sided Hypothesis Test of the Population Mean (σ_x known)

Conditions:

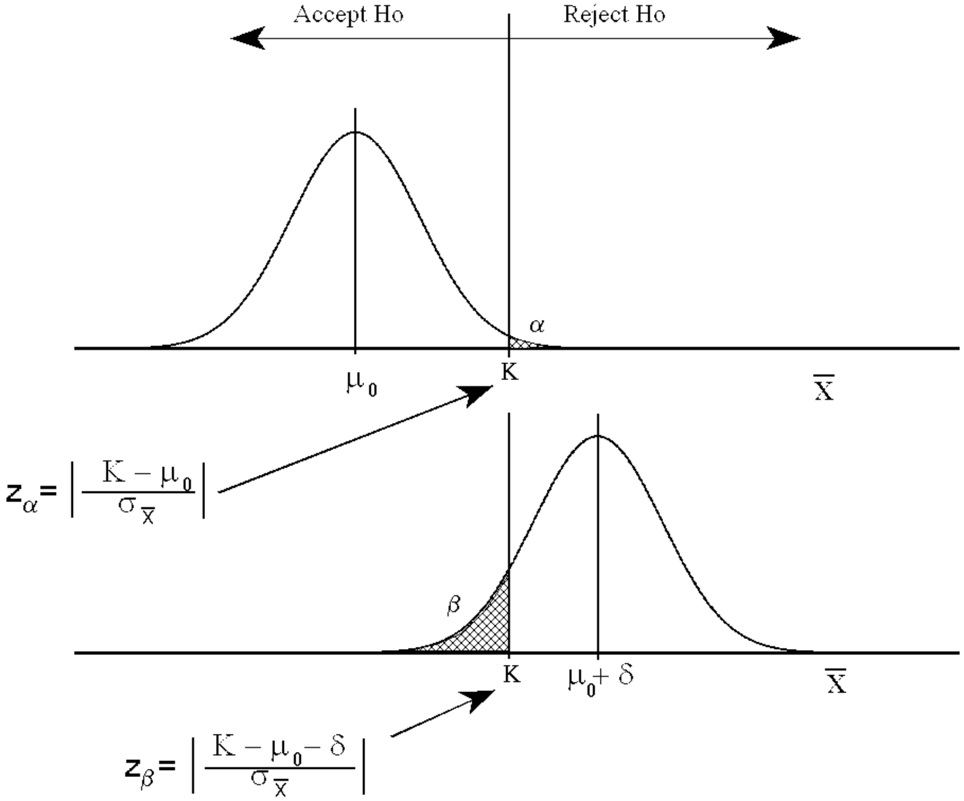
- σ_x is known
- x is normally distributed.

Hypotheses: $H_0 : \mu = \mu_0$ vs. $H_A : \mu > \mu_0$ or alternatively, $H_0 : \delta = 0$ vs. $H_A : \delta > 0$ where $\delta = \mu - \mu_0$.

Sample Size: The sample size required to obtain power $P = 1 - \beta$ for a shift from $\mu = \mu_0$ to $\mu = \mu_0 + \delta$ is given by:

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

where z_α and z_β are both positive.



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

$$K = \mu_0 + \delta\left(\frac{z_\alpha}{z_\alpha + z_\beta}\right)$$

Example: An experiment will be performed to determine if the burst pressure of a small pressure vessel is 60psi or if the burst pressure is greater than 60psi. The standard deviation of burst pressure is known to be 5psi and the experiment should reject $H_0 : \mu = 60$ with 90% probability if $\mu = 63$. Determine the sample size and acceptance condition for the experiment. The distribution of x is normal and use $\alpha = 0.05$.

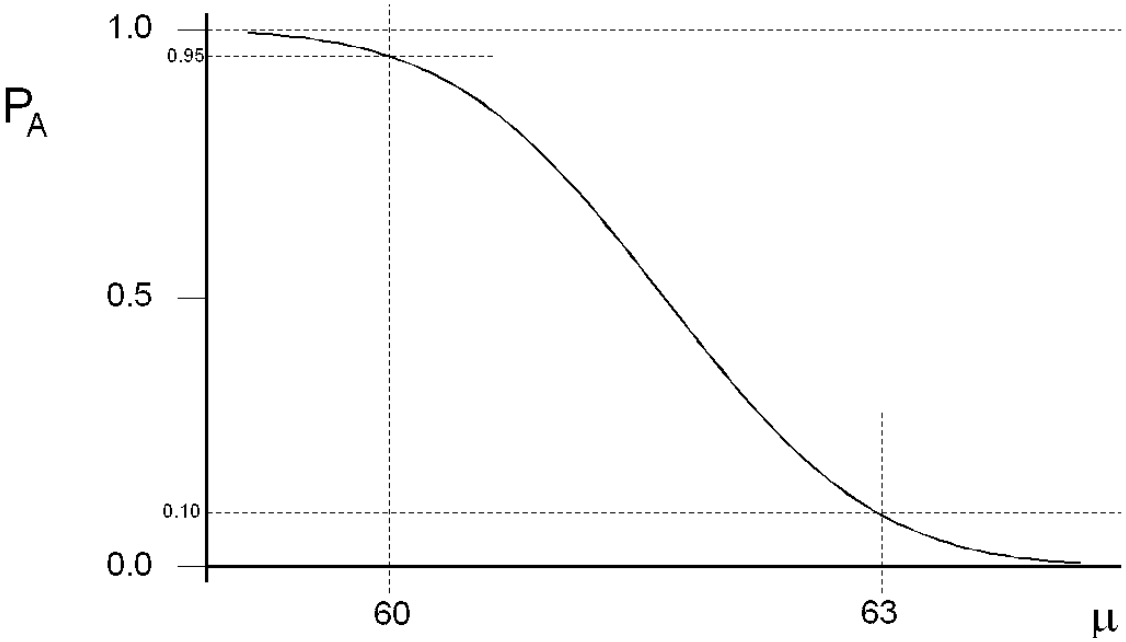
Solution: The hypotheses to be tested are $H_0 : \mu = 60$ vs. $H_A : \mu > 60$. The power of the experiment to reject H_0 when $\mu = 63$ or $\delta = 3$ is $P = 1 - \beta = 0.90$ so $\beta = 0.10$. The sample size is given by:

$$\begin{aligned} n &= \left(\frac{(z_{0.05}+z_{0.10})\sigma_x}{\delta}\right)^2 \\ &= \left(\frac{(1.645+1.282)5}{3}\right)^2 \\ &= 24 \end{aligned}$$

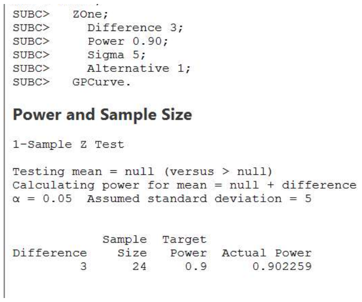
The critical accept/reject value of $\bar{x}$ is given by:

$$\begin{aligned} K &= \mu_0 + \delta\left(\frac{z_{0.05}}{z_{0.05}+z_{0.10}}\right) \\ &= 60 + 3\left(\frac{1.645}{1.645+1.282}\right) \\ &= 61.69 \end{aligned}$$

The following graph shows the OC curve for the sampling plan:



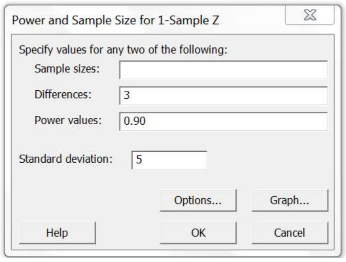
Using MINITAB **Stat**> **Power and Sample Size**> **1-Sample Z**:



```
SUBC>   ZOne;
SUBC>     Difference 3;
SUBC>     Power 0.90;
SUBC>     Sigma 5;
SUBC>     Alternative 1;
SUBC>   GPCurve.
```

Power and Sample Size

```
1-Sample Z Test

Testing mean = null (versus > null)
Calculating power for mean = null + difference
α = 0.05  Assumed standard deviation = 5


            Sample  Target
Difference    Size   Power  Actual Power
         3      24     0.9      0.902259
```



Power and Sample Size for 1-Sample Z

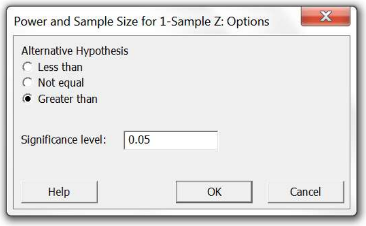
Specify values for any two of the following:

Sample sizes:

Differences: 3

Power values: 0.90

Standard deviation: 5



Power and Sample Size for 1-Sample Z: Options

Alternative Hypothesis

○ Less than

○ Not equal

● Greater than

Significance level: 0.05

Sample Size for a Two-Sided Hypothesis Test of the Population Mean (σ_x known)

Conditions:

- σ_x is known
- x is normally distributed.

Hypotheses: $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ or alternatively, $H_0 : \delta = 0$ vs. $H_A : \delta \neq 0$ where $\delta = |\mu_0 - \mu|$.

Sample Size: The sample size required to reject $H_0 : \mu = \mu_0$ with probability $P = 1 - \beta$ for a shift from $\mu = \mu_0$ to $\mu = \mu_0 \pm \delta$ is given by:

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

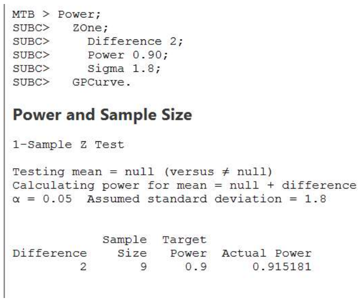
where $z_{\alpha/2}$ and z_β are both positive.

Example: Determine the sample size required to detect a shift from $\mu = 30$ to $\mu = 30 \pm 2$ with probability $P = 0.90$. Use $\alpha = 0.05$. The population standard deviation is $\sigma_x = 1.8$ and the distribution of x is Φ.

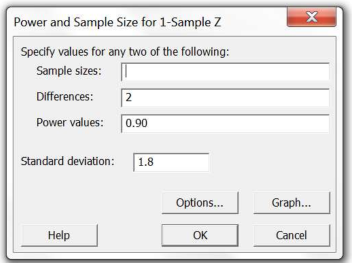
Solution: The hypotheses being tested are $H_0 : \mu = 30$ vs. $H_A : \mu \neq 30$. The size of the shift that we want to detect is $\delta = 2$ and we have $\sigma = 1.8$. Since $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_\beta = z_{0.10} = 1.28$ the sample size required for the test is:

$$\begin{aligned} n &= \left(\frac{(z_{\alpha/2}+z_\beta)\sigma_x}{\delta}\right)^2 \\ &= \left(\frac{(1.96+1.28)1.8}{2}\right)^2 \\ &= 8.5 \rightarrow 9 \end{aligned}$$

Using MINITAB **Stat**> **Power and Sample Size**> **1-Sample Z**:



```
MTB > Power;
SUBC>   ZOne;
SUBC>     Difference 2;
SUBC>     Power 0.90;
SUBC>     Sigma 1.8;
SUBC>   GPCurve.

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus ≠ null)
Calculating power for mean = null + difference
α = 0.05  Assumed standard deviation = 1.8


            Sample  Target
Difference    Size   Power  Actual Power
         2       9     0.9      0.915181
```



Power and Sample Size for 1-Sample Z

Specify values for any two of the following:

Sample sizes:

Differences: 2

Power values: 0.90

Standard deviation: 1.8

Sample Size for Hypothesis Tests for the Difference Between Two Population Means

Conditions:

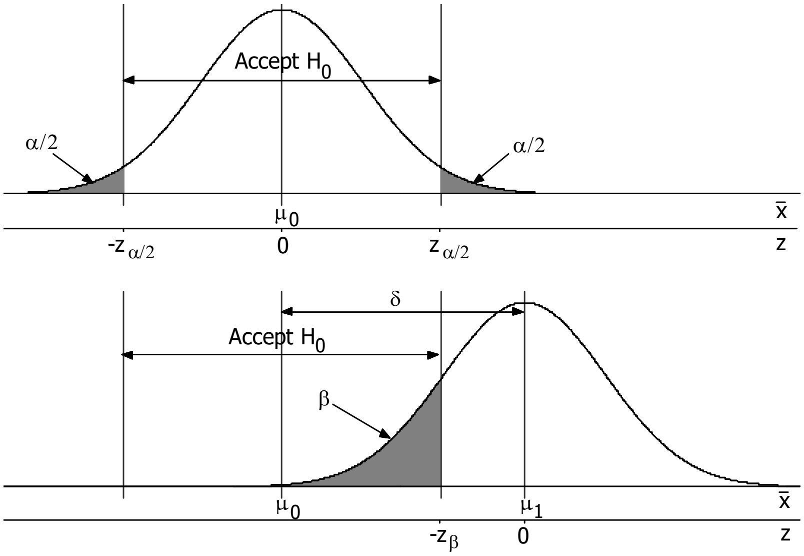
- σ_1 and σ_2 are both known and $\sigma_1 = \sigma_2$
- x_1 and x_2 are normally distributed

Hypotheses: $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ or alternatively, $H_0 : \delta = 0$ vs. $H_A : \delta \neq 0$ where $\delta = |\mu_1 - \mu_2|$.

Sample Size: The sample size required to reject H_0 with probability $P = 1 - \beta$ for a difference between the means of $|\mu_1 - \mu_2| = \delta$ is given by:

$$n_1 = n_2 = 2\left(\frac{(z_{\alpha/2} + z_\beta)\sigma_x}{\delta}\right)^2$$

where $z_{\alpha/2}$ and z_β are both positive. For the one-sided tests replace $z_{\alpha/2}$ with z_α.

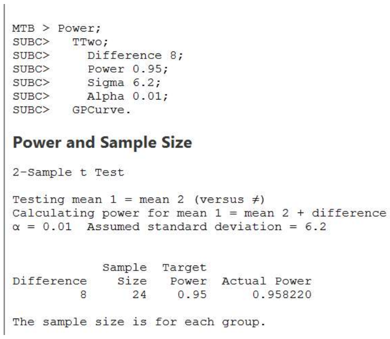


Example: Determine the common sample sizes required to detect a difference between two population means of $|\mu_1 - \mu_2| = \delta = 8$ with probability $P = 0.95$. Use $\alpha = 0.01$. The population standard deviation is $\sigma_x = 6.2$ and the distribution of x is Φ.

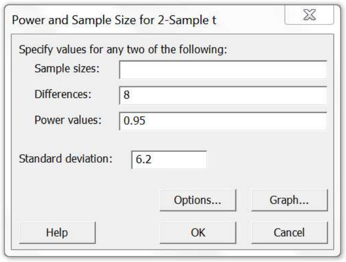
Solution: The hypotheses to be tested are $H_0 : \delta = 0$ vs. $H_A : \delta \neq 0$. We want to detect a difference between the two means of $\delta = 8$ with probability $P = 0.95$ so we have $\beta = 1 - P = 0.05$ so $z_\beta = z_{0.05} = 1.645$. For the two-tailed test we need $z_{\alpha/2} = z_{0.005} = 2.575$ so the required sample size is:

$$\begin{aligned} n_1 = n_2 &= 2\left(\frac{(z_{\alpha/2}+z_\beta)\sigma_x}{\delta}\right)^2 \\ &= 2\left(\frac{(2.575+1.645)6.2}{8}\right)^2 \\ &= 21.4 \rightarrow 22 \end{aligned}$$

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



```
MTB > Power;
SUBC>   TTwo;
SUBC>     Difference 8;
SUBC>     Power 0.95;
SUBC>     Sigma 6.2;
SUBC>     Alpha 0.01;
SUBC>   GPCurve.

Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus ≠)
Calculating power for mean 1 = mean 2 + difference
α = 0.01  Assumed standard deviation = 6.2


            Sample  Target
Difference    Size   Power  Actual Power
         8      24    0.95      0.958220

The sample size is for each group.
```



Power and Sample Size for 2-Sample t

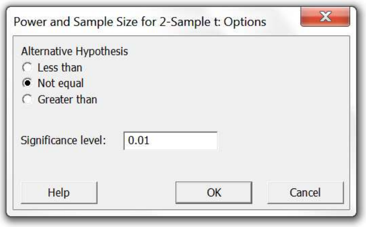
Specify values for any two of the following:

Sample sizes:

Differences: 8

Power values: 0.95

Standard deviation: 6.2



Power and Sample Size for 2-Sample t: Options

Alternative Hypothesis

○ Less than

● Not equal

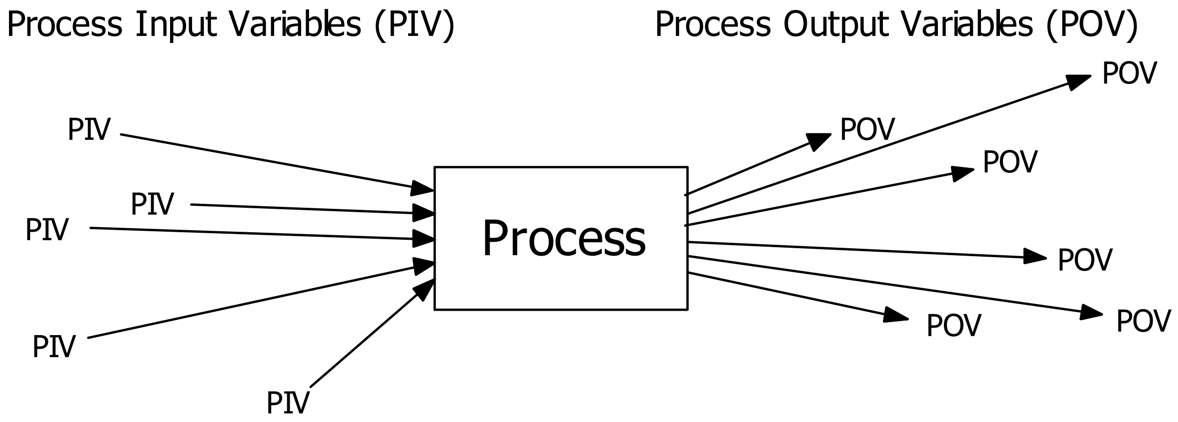
○ Greater than

Significance level: 0.01

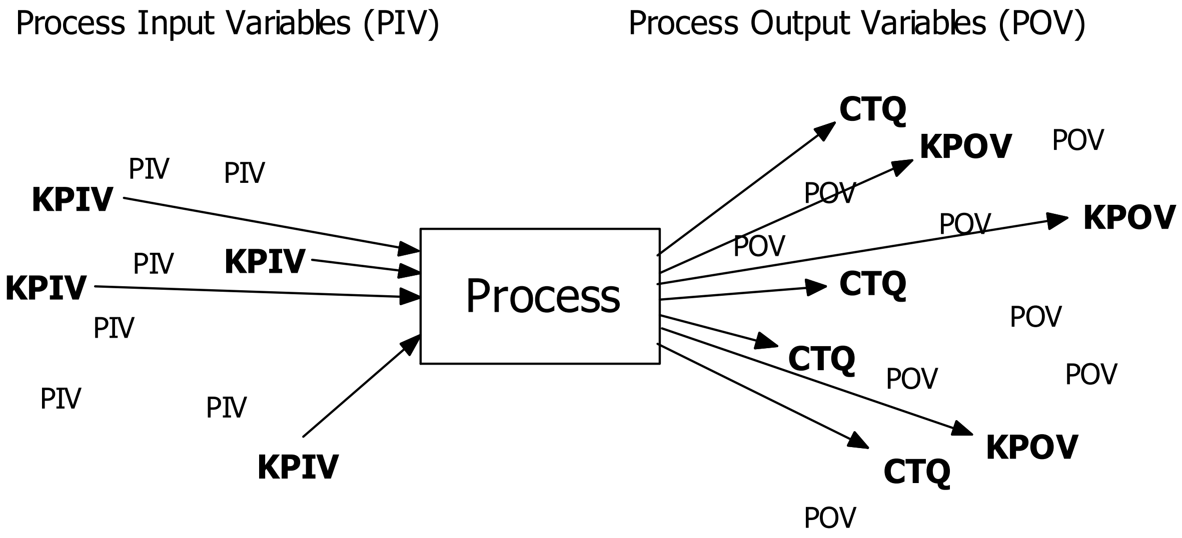
Chapter 4: The Language of DOE

Input-Process-Output Diagrams

Use an input-process-output (IPO) diagram to catalog all of the possible input and output variables of a process:



The goal is to manage the KPIVs so that all of the requirements of the CTQs and KPOVs are satisfied:

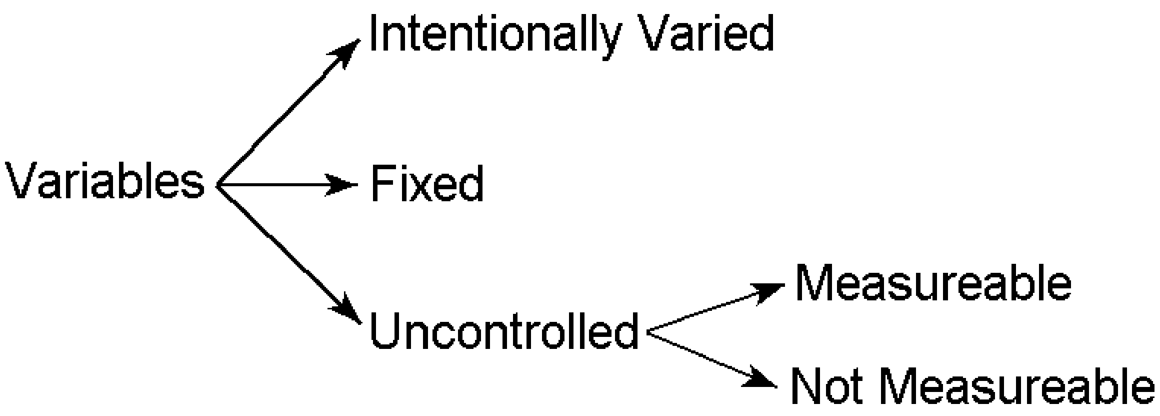


KPIV = Key Process Input Variable

CTQ = Critical To Quality
KPOV = Key Process Output Variable

"The novice sees many possibilities. The expert sees few." - Shunryu Suzuki

Disposition of Design Variables in an Experiment



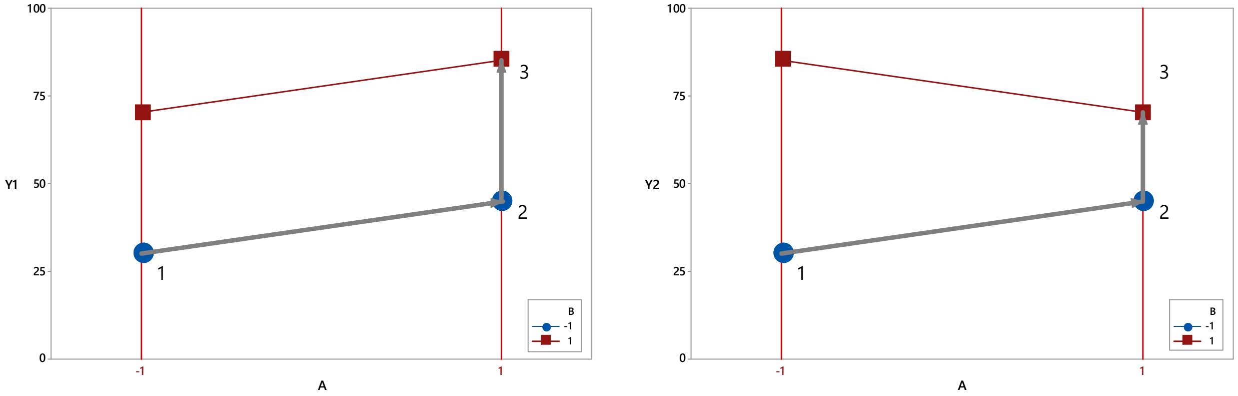
Variable Types

- Quantitative variables
  - Require a valid measurement scale
- Qualitative variables
  - Fixed: All levels are known and identified.
  - Random: Levels are random sample of many possible levels.
- We will limit our considerations to quantitative response variables.
- Design (i.e. input) variables will be both qualitative and quantitative.

Why Is DOE Necessary?

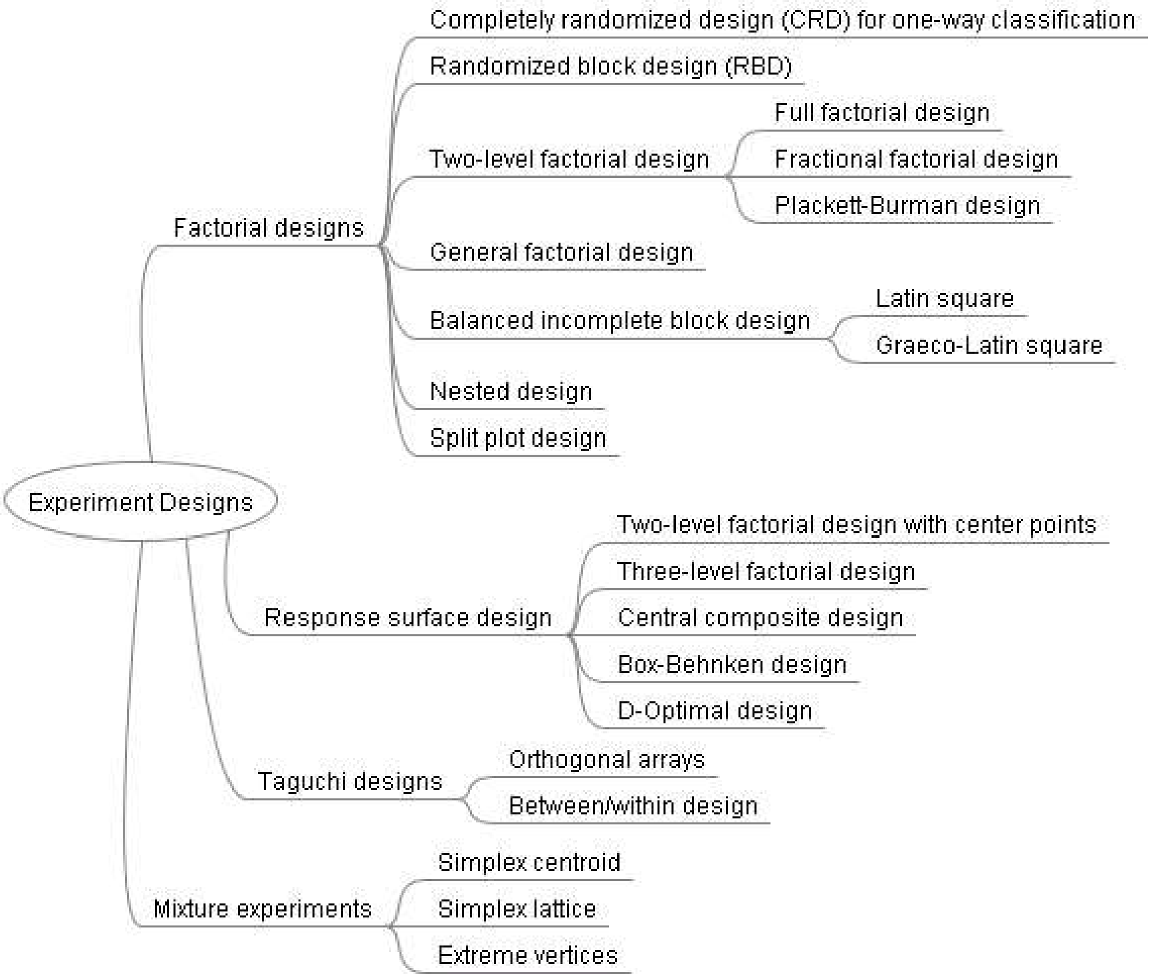
DOE allows the simultaneous investigation of the effect of several variables on a response in a cost effective manner. DOE is superior to the traditional one-variable-at-a-time method (OVAT).

Example: Find the values of x_1 and x_2 that maximize the response by the OVAT method. OVAT fails in the second case because there is an *interaction* between variables A and B that the OVAT method cannot resolve.



Types of Experiments

- Screening Experiments
  - Good first experiment
  - Can consider many variables
  - Pareto mode: Identify the few important variables among the many
  - Usually only two levels of each variable
  - Relatively few runs
  - Limited if any ability to identify interactions
  - Risky
- Factorial and Response Surface Experiments
  - Good follow-up experiment to a screening experiment
  - Fewer variables - generally the most important ones
  - Often three or more levels of each variable
  - Provide a more complex model for the process



Relationship Between the Familes of Design Experiments

- Projects or programs to study a complicated process usually require more than one experiment:
  - a series of sequential experiments (see below)
  - iterative experiments to clarify missed variables, poor variable level choices, procedural errors, and other oversights
- A procedure for sequential experiments - progressing from simple to complex models:
  1. Start from the present understanding of the process.
  2. Screening experiment - Distinguish which of many variables are the most important:

     $$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots$$

  3. Factorial experiment - Quantify variable effects, two-factor interactions, and maybe check for curvature:

     $$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_{12} x_{12} + \cdots + b_{**} x_*^2$$

  4. Response surface design - Add quadratic terms to account for curvature:

     $$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_{12} x_{12} + \cdots + b_{11} x_1^2 + b_{22} x_2^2 + \cdots$$

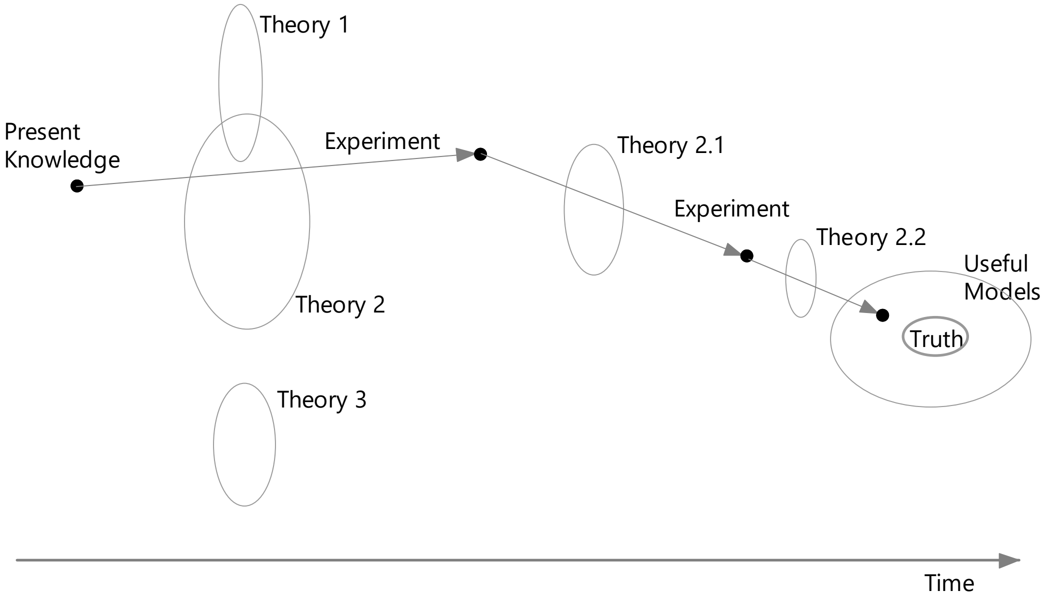
  5. Arrive at a useful model.

Types of Models

- Model with a qualitative PIV:
  - Requires that the mean of each level be specified, e.g. five levels require specification of $\bar{x}_1, \bar{x}_2, ..., \bar{x}_5$ to estimate $\mu_1, \mu_2, ..., \mu_5$.
  - Analysis is by ANOVA.
- Model with a quantitative PIV:
  - Requires mathematical expression of $y = f(x)$ in the form of an equation which can be linear, quadratic, etc.
  - Analysis is by regression.
- Types of models:
  - First principles model - based on first principles of physics, mechanics, chemistry, ...
  - Empirical - absent knowledge of a first principles model use a Taylor expansion:

    $$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_{12} x_{12} + \cdots + b_{11} x_1^2 + b_{22} x_2^2 + \cdots$$

  - Even when the form of a first principles model is unknown, first principles should still be used to inform the empirical model.
- "All models are wrong. Some are useful." George Box



What is a Model?

Data contain information and noise. A model is a concise mathematical way of describing the information content of the data, however; any model must be associated with a corresponding error statement that describe the noise:

$$Data = Model + Error\ Statement$$

When you are trying to communicate information to someone you can either give them all of the data and let them draw their own conclusions or state a model for the data and describe the discrepancies from the model.

The description of the errors must include: 1) the shape of the distribution of errors and 2) the size of the errors.

Model for a Single Set of Measurement Values

Example: 5000 normally distributed observations (x_i) have a mean $\bar{x} = 42$ and a standard deviation of $s = 2.3$. Identify the data, model, and error in this situation.

Solution: The data are the 5000 observations x_i. The model is $\hat{x}_i = \bar{x}$. The errors are normally distributed about $\bar{x}$ with standard deviation $s = 2.3$.

$$\underbrace{(x_1, x_2, \ldots, x_{5000})}_{Data} = \underbrace{\bar{x}}_{Model} \quad \underbrace{and\ \Phi(\epsilon_i; 0, s)}_{Error\ Statement}$$

Model for a Set of Paired (x,y) Quantitative Observations

Example: 200 paired observations (x_i, y_i) are collected. A line is fitted to the data and the resulting fit is $\hat{y}_i = 80 - 5x_i$. The points are scattered randomly above and below the fitted line in a normal distribution with a standard error of $s_\epsilon = 2.3$. Identify the data, model, and error in this situation.

Solution: The data are the 200 observations (x_i, y_i). The model is $\hat{y}_i = 80 - 5x_i$. The errors are normally distributed about the fitted line with standard deviation $s_\epsilon = 2.3$.

$$\underbrace{\{(x_1, y_1), (x_2, y_2), \ldots, (x_{200}, y_{200})\}}_{Data} = \underbrace{80 - 5x_i}_{Model} \quad \underbrace{and\ \Phi(\epsilon_i; 0, 2.3)}_{Error\ Statement}$$

Model for a One-way Classification

Example: Forty measurements are taken from five different lots of material. The lot means are $520, 489, 515, 506,$ and 496. The errors within the lots are normally distributed with a standard error of 20. Identify the data, the model, and the error.

Solution: The data are the 40 observations taken from 5 different populations. The model is provided by the 5 means: $520, 489, 515, 506,$ and 496. The error statement is that the errors are normally distributed about the lot means with a standard deviation of $s_\epsilon = 20$.

$$\begin{aligned}
(x_{11}, x_{12}, \ldots, x_{18}) &= 520 \\
(x_{21}, x_{22}, \ldots, x_{28}) &= 489 \\
(x_{31}, x_{32}, \ldots, x_{38}) &= 515 \\
(x_{41}, x_{42}, \ldots, x_{48}) &= 506 \\
\underbrace{(x_{51}, x_{52}, \ldots, x_{58})}_{Data} &= \underbrace{496}_{Model} \quad \underbrace{and\ \Phi(\epsilon_i; 0, 20)}_{Error\ Statement}
\end{aligned}$$

Selection of Study (PIV) Variable Levels

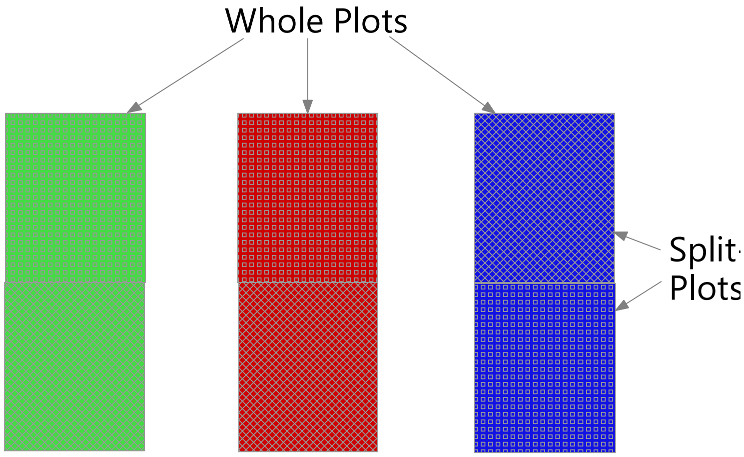
- Variables may be qualitative or quantitative
  - Each study variable must have at least two levels
- Qualitative variables, e.g. operators, material lots, ...
  - Fixed levels
    - Finite number of levels and all are available
    - Goal is to estimate biases between levels
  - Random levels
    - Too many levels to include them all in the experiment
    - Limited to a random sample
    - Goal is to estimate the standard deviation of biases between levels
- Quantitative variables, e.g. temperature, pressure, dimension, ...
  - Two levels (e.g. low and high) is sufficient to quantify main effects and two-factor interactions
  - Three or more levels are required to resolve quadratic terms
  - More than three levels are required to resolve higher order terms but we usually don't have to go that far
  - Spacing:
    - Too close together and you won't see an effect
    - Too far apart and one or both levels may not work
    - Too far apart and an approximately linear relationship can go quadratic or worse

Nested Variables

- The levels of a variable are unique within one level of another variable.
- Examples:
  - Operators within shifts
  - Heads within machines
  - Cavities within a multi-cavity mold
  - Subsamples from samples from cups from totes from lots from a large production run of a dry powder

Split Plots

- The name comes from agricultural experiments, where different hard-to-change treatments were applied to large areas of a field (plots) and different easy-to-change treatments were applied to smaller areas within plots (sub- or split-plots).
- A split-plot design is a hybrid or cross of two experiment designs, one design involving hard-to-change (HTC) variables and a second design involving easy-to-change (ETC) variables.



What is an Experiment Design?

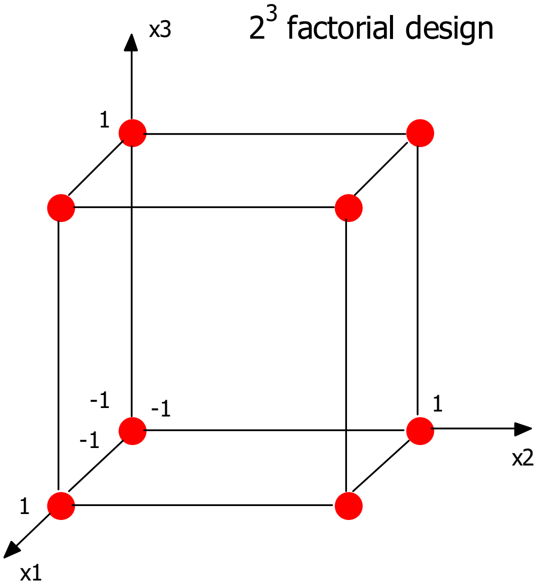
- The variables matrix defines the levels of the design variables:

Level	x_1:Batch Size	x_2:Resin	x_3:Mixing Time
-	50cc	A	1 minute
+	150cc	B	3 minutes

- The experiment design matrix defines the combination of levels used in the experiment:

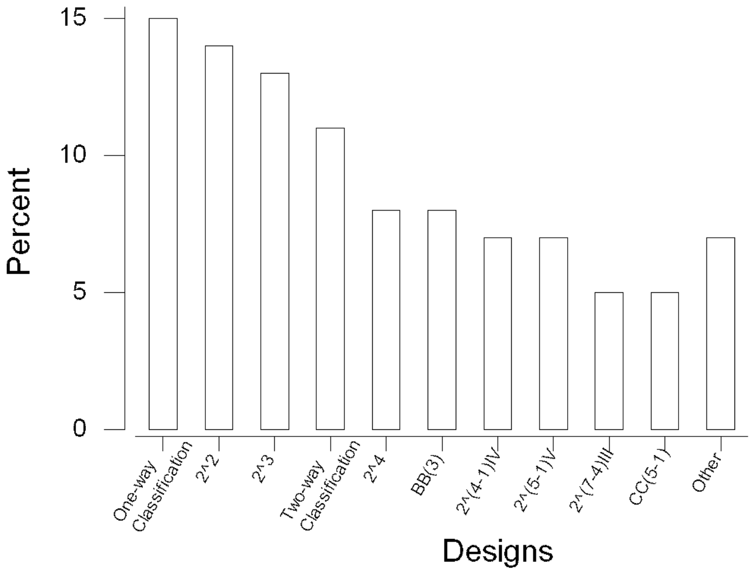
Run	x_1:Batch Size	x_2:Resin	x_3:Mixing Time
1	-	-	-
2	-	-	+
3	-	+	-
4	-	+	+
5	+	-	-
6	+	-	+
7	+	+	-
8	+	+	+

This experiment design is called a 2^3 design because there are three variables, each at two levels, so there are $2^3 = 8$ unique experimental runs.



- The purpose of breaking the experiment design up into two matrices, the variables matrix and the design matrix, is to distinguish between the sources of expertise required to produce them. The variables matrix requires substantial information that can only come from the process owner whereas the design matrix can be chosen by anyone skilled in DOE methods.

Most Experiments Use Just a Few Designs



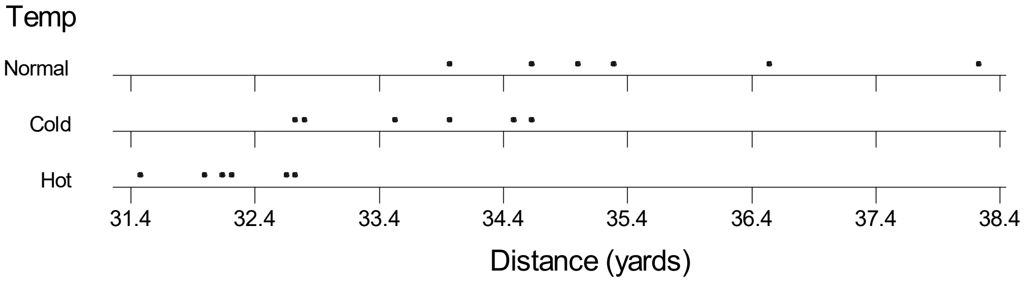
Other Issues

- **Extra and Missing Runs** - Avoid building extra runs or losing runs from the experiment. Extra and missing runs unbalance the experiment design and cause undesireable correlations between terms in the model that compromise its integrity. Methods to deal with such problems will be addressed later.
- **Randomization** - If claims are to be made about differences between the levels of a variable, then the run order of the levels in the experiment must be randomized. Randomization protects against the effects of unidentified or "lurking" variables.
- **Blocking** - If the run order of the levels of a variable is not randomized then that variable is a blocking variable. This is useful for isolating variation between blocks but claims can not be made about the true cause of differences between the blocks. **Variation due to uncontrolled sources should be homogeneous within blocks but can be heterogeneous between blocks**.
- **Repetition** - Consecutive observations made under the same experimental conditions. Repetitions are usually averaged and treated as a single observation so they are oftern of negligible value.
- **Replication** - Experimental runs made under the same settings of the study variables but at different times. Replicates carry more information than repetitions. The number of replicates is an important factor in determining the sensitivity of the experiment.
- **Confounding** - Two design variables are confounded if they predict each other, i.e. if their values are locked together in some fixed pattern. The effects of confounded variables cannot be separated. Confounding should be avoided (best practice) or managed (a compromise).

Case Study

(http://youth.net/nsrc/sci/sci059.html, with permission from John Strang.) A student performed a science fair project to study the distance that golf balls traveled as a function of golf ball temperature. To standardize the process of hitting the golf balls, he built a machine to hit balls using a five iron, a clay pigeon launcher, a piece of plywood, two sawhorses, and some duct tape. The experiment was performed using three sets of six Maxfli golf balls. One set of golf balls was placed in hot water held at $66C$ for 10 minutes just before they were hit, another set was stored in a freezer at $-12C$ overnight, and the last set was held at ambient temperature ($23C$). The distances in yards that the golf balls traveled are shown in the table below but the order used to collect the observations was not reported. Create dotplots of the data and interpret the differences between the three treatment means assuming that the order of the observations was random. How does your interpretation change if the observations were collected in the order shown - all of the hot trials, all of the cold trials, and finally all of the ambient temperature trials?

	Trial					
Temp	1	2	3	4	5	6
$66C$	31.50	32.10	32.18	32.63	32.70	32.00
$-12C$	32.70	32.78	33.53	33.98	34.64	34.50
$23C$	33.98	34.65	34.98	35.30	36.53	38.20



Golf Ball Distance vs. Temperature

General Procedure for Experimentation

The following 11 step procedure outlines all of the steps involved in planning, executing, analyzing, and reporting an experiment ...

1. Prepare a cause and effect analysis of all of the process inputs (variables) and outputs (responses).
2. Document the process using written procedures or flow charts.
3. Write a detailed problem statement.
4. Perform preliminary experimentation.
5. Design the experiment.
6. Determine the number of replicates and the blocking and randomization plans.
7. Run the experiment.
8. Perform the statistical analysis of the experimental data.
9. Interpret the statistical analysis.
10. Perform a confirmation experiment.
11. Report the results of the experiment.

General Procedure for Experimentation

1. Input-Process-Output (IPO) Diagram
    a. Catalog all of the input variables: methods, manpower, machines, material, and environment.
    b. Catalog all of the possible responses.
    c. Make the catalogs exhaustive!
    d. Brainstorm everything.
    e. Reevaluate and revise this list regularly!
2. Document the Process to be Studied
    a. Review or cite the theory of the process.
    b. Review the process flow charts and written procedures.
    c. Review calibration and gage error study results for all measurement variables (inputs and outputs).
    d. Review process capability studies, SPC charts, and process logs.
    e. Identify workmanship examples.
    f. Talk to the operators or technicians who do the work.
    g. Identify training opportunities.
    h. Get general agreement on all steps of the process.
3. Write a Detailed Problem Statement or Protocol Document
    a. Identify the response(s) to be studied.
    b. Identify the design variables.
        i. Variables for active experimentation.
        ii. Variables to be held fixed.
        iii. Variables that cannot be controlled.
    c. Identify possible interactions between variables.
    d. Estimate the repeatability and reproducibility.
    e. Cite evidence of gage capability.
    f. Cite evidence that the process is in control.
    g. Identify assumptions.
    h. State the goals and limitations of the experiment.
    i. Estimate the time and materials required.
    j. Identify knowledge gaps.
4. Preliminary Experimentation
    a. Used to resolve knowledge gaps.
    b. Determine nature of and levels for input variables:
        i. Quantitative or qualitative?
        ii. Fixed or random?
        iii. Too narrow and you won’t see an effect.
        iv. Too wide and you may lose runs or get curvature.
    c. Use no more then 15% of your resources.
    d. Refine the experimental procedure.
    e. Confirm that the process is in control.
    f. Confirm that all equipment is operating correctly and has been maintained.

5. Design the Experiment
    a. Assumption: The intended model and analysis method for $y = f(x_1, x_2, \cdots)$ are known.
    b. Select an experiment design:
        i. Screening experiment.
        ii. Experiment to resolve main effects and interactions.
        iii. Response surface experiments.
    c. Consider opportunities to add a variable.
    d. Identify and evaluate the merits of alternative designs.
    e. Plan to use no more than about 70% of your resources.
6. Replicates, Randomization, and Blocking
    a. Determine the number of replicates.
    b. Build large experiments in blocks.
    c. You MUST randomize. Failure to randomize may lead to incorrect conclusions and leaves your claims open to challenge.
    d. Randomize study variables within blocks.
    e. Validate your randomization plan.
    f. Design data collection forms.
7. Conduct the Experiment
    a. Make sure all critical personnel, materials, and equipment are available and functional.
    b. Record all of the data.
    c. Note any special occurrences.
    d. If things go wrong decide whether to postpone the experiment or whether to revise the experiment design and/or procedure.
8. Analyze the Data
    a. Confirm the accuracy of the data.
    b. Graph the data.
    c. Run the ANOVA or regression.
    d. Check assumptions:
        i. Orthogonality
        ii. Equality of variances
        iii. Normality of residuals
        iv. Independence
        v. Check for lack of fit
    e. Refine the model using Occam's Razor.
    f. Determine the model standard error and R-squared.
    g. Consider alternative models.
9. Interpret the Results
    a. Develop a predictive model for the response.
    b. Does the model make sense?
    c. Select the optimum variable levels.
    d. Don't extrapolate outside the range of experimentation.
    e. Plan a follow-up experiment to resolve ambiguities.
10. Perform a Confirmation Experiment
    a. Validate the model by showing that you can achieve the same result again.
    b. Use the remaining 10% of your resources.
    c. Don't report any results until after the confirmation experiment is complete.
11. Document the Results
    a. Keep all of the original records and notes.
    b. Write the formal report.
    c. Know your audience.

Who Is Involved? What Are Their Responsibilities?

Activity	Project Leader	Operators	Technicians	Design Engineer	Process Engineer	Manager/ Customer	Statistical Specialist
1. Cause and Effect Analysis	√	√	√	√	√	√	
2. Document the Process	√	√	√	√	√		
3. Problem Statement	√	Review	Review	Review	Review	Review	Review
4. Preliminary Experiment	√	√	√	√	√		
5. Design the Experiment	√						Support
6. Randomization Plan	√						Support
7. Run the Experiment	√	√	√	√	√		
8. Analyze the Data	√						Support
9. Interpret the Model	√						Support
10. Confirmation Experiment	√	√	√				
11. Report the Results	√			Review	Review	Review	

Organization Culture and Infrastructure for Experiments

- Organizations must develop the culture and infrastructure necessary to run successful programs of experiments.
- Some companies/organizations have a mature environment for adminstrating experiments that permits a relatively informal experiment management system.
- Other companies/organizations may demand (by choice) or require (highly regulated industry, contract research lab, consulting, SBIR or STTR grant application, etc.) a more structured approach. The key document in the planning and execution of an experiment it this environment is the *experiment protocol* document.
- Components of an Experiment Protocol
  - Administrative Information: title, author, date, etc.
  - Introduction
  - Experiment design
  - Sample size, blocking, randomization plan
  - Experimental procedure
  - Data recording
  - Statistical analysis
  - Report format

Why Experiments Go Bad

- “An expert is a person who has made all the mistakes that can be made in a very narrow field." - Niels Bohr
- "The 9/11 Commission identified four types of systemic failures ..., failures of policy, capabilities, and management. The most important category of failure was failure of imagination." - Nate Silver, The Signal and the Noise
- There are known knowns; there are things that we know we know. We also know that there are known unknowns; that is to say, we know there are some things that we do not know. But there are also unknown unknowns; there are things we do not know we don’t know." - Donald Rumsfeld

Why Experiments Go Bad

- Inexperienced experimenter
- The presence of the experimenter changes the process
- Failure to identify an important variable
- Picked the wrong variables for the experiment
- Failure to hold a known variable fixed
- Failure to record the value of a known but uncontrollable variable
- Poor understanding of the process and procedures
- Failure to consult the operators and technicians
- Failure to anticipate or plan for significant effect, e.g. interaction or quadratic term
- Failure to recognize all of the responses
- Inadequate R&R to measure the response
- Inadequate R&R for a quantitative predictor
- Failure to account for noise in a predictor intended to have fixed levels
- Used incorrect variable level
- Failure to do any or enough preliminary experimentation
- Exhausted resources and patience with too much preliminary experimentation
- Picked variable levels too close together
- Picked variable levels too far apart
- Wrong experiment design
- One experiment instead of several smaller ones
- Several small experiments instead of a single larger one
- Not enough replicates
- Repetitions instead of replicates
- Failure to randomize
- Randomization plan ignored by those running the experiment
- Failure to record the actual run order
- Failure to block the experiment to control the effects of lurking variables
- Failure to run controls
- Critical person missing when experiment is run
- Failure to record all of the data
- Failure to maintain part identity
- Unanticipated process change during experiment
- Equipment not properly maintained
- Failure to complete the experiment in the allotted time (e.g. before a shift change)
- Failure to note special occurrences
- Wrong statistical analysis
- Failure to check assumptions (normality, equality of variances, lack of fit, ...)
- Failure to specify the model correctly in the analysis software
- Mistreatment of lost experimental runs
- Failure to refine the model
- Misinterpretation of results
- Extrapolation outside of experimental boundaries
- Failure to perform a confirmation experiment
- Inadequate resources to build a confirmation experiment
- Inadequate documentation of the results
- Inappropriate presentation of the results for the audience

Chapter 5: Experiments for One-way Classifications

The Purpose of ANOVA

- The purpose of ANOVA is to determine if one or more pairs of treatment means among three or more treatments are different from the others:

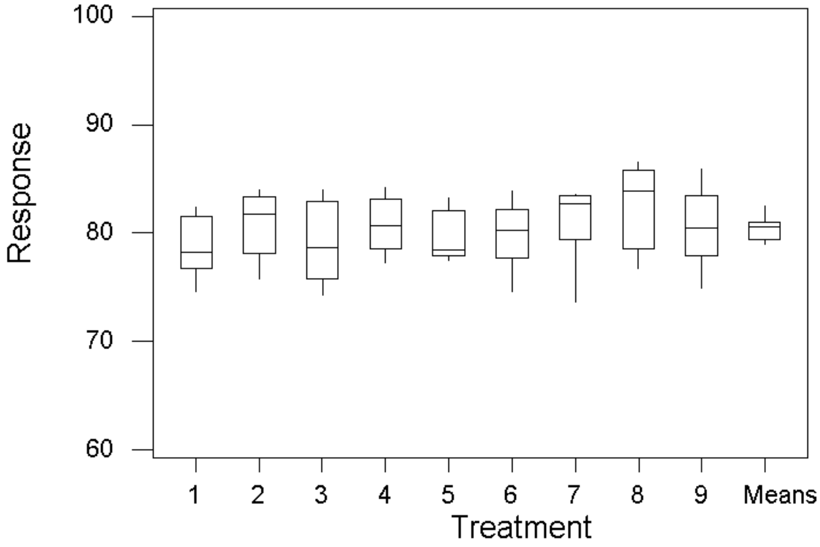
$$H_0 : \mu_i = \mu_j \text{ for all possible pairs}$$

$$H_A : \mu_i \neq \mu_j \text{ for at least one pair}$$

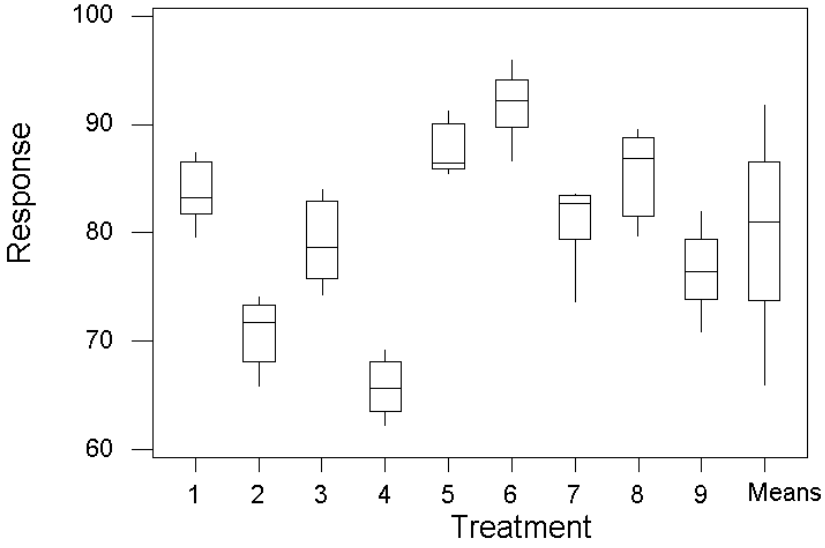
- ANOVA doesn't indicate which pairs of means are different, so follow-up multiple comparison test (MCT) methods are used after ANOVA.

The Graphical Approach to ANOVA

If H_0 is true, then $\sigma_{\bar{y}} = \sigma_y/\sqrt{n}$:



If H_0 is false, then $\sigma_{\bar{y}} \gg \sigma_y/\sqrt{n}$:

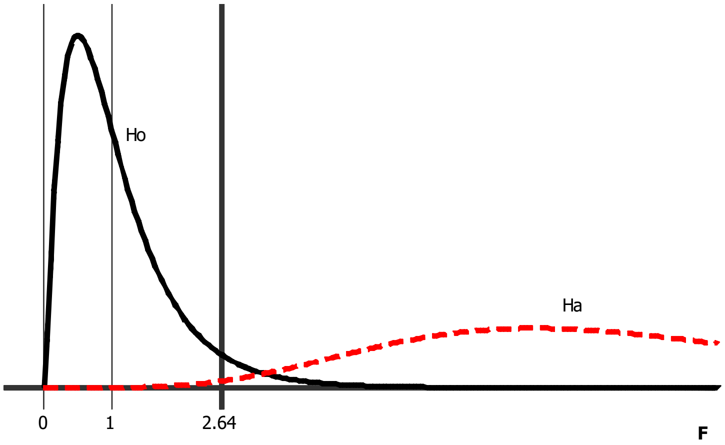


The Key to ANOVA is an F Test

The ANOVA F test compares two independent estimates of the population variance determined from the variation between treatments $\left(\widehat{\sigma}_{\bar{y}}^2\right)$ and the variation within treatments $\left(\widehat{\sigma}_{\epsilon}^2\right)$. If $H_0 : \mu_i = \mu_j$ for all i,j is true, then by the central limit theorem $\sigma_y^2 = n\sigma_{\bar{y}}^2$ so

$$F = \frac{\widehat{\sigma}_y^2}{\widehat{\sigma}_y^2} = \frac{ns_{\bar{y}}^2}{s_{\epsilon}^2}$$

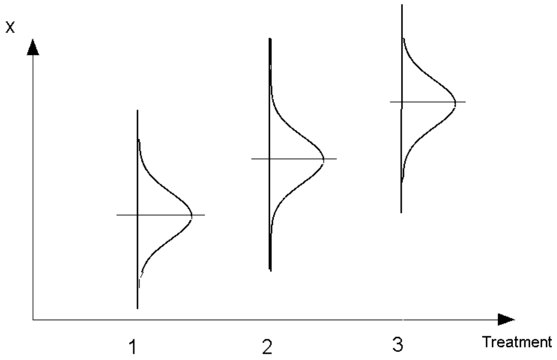
follows the F distribution. When H_0 is true, then $E(F) = 1$. When $H_0 : \mu_i = \mu_j$ is not true then $E(F) > 1$.



ANOVA Assumptions

ANOVA requires that the following assumptions are met:

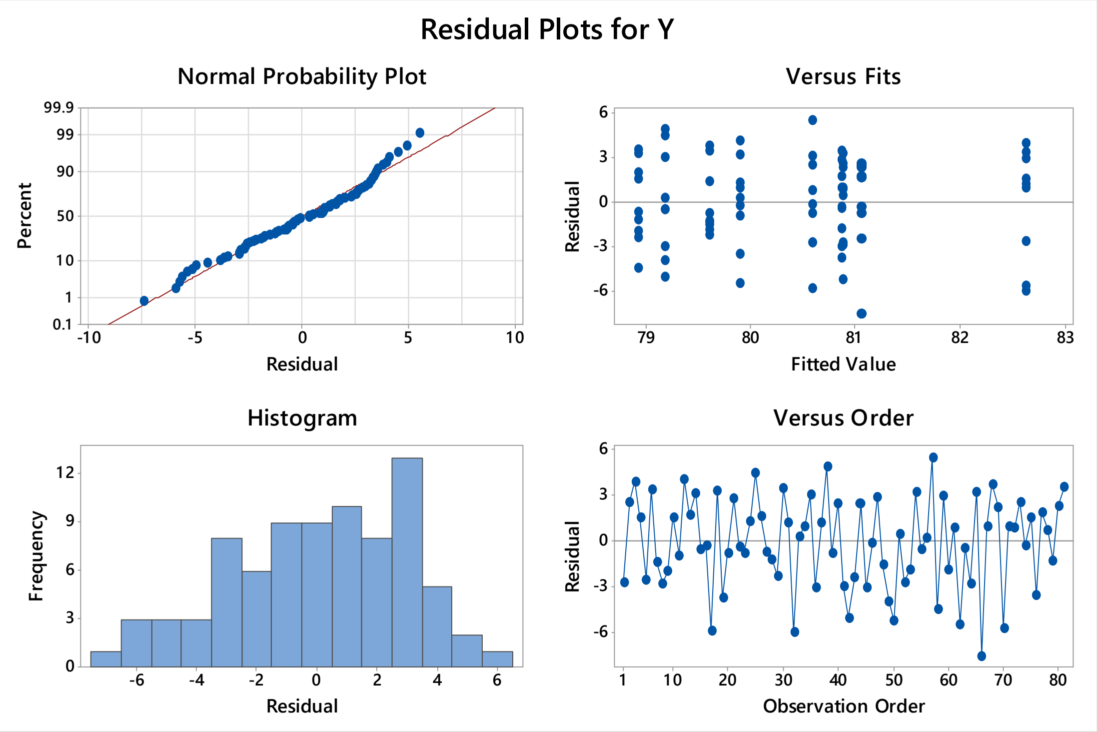
- The k populations being sampled are normally distributed.
- The k populations being sampled have equal variances, i.e. are homoscedastic.
- The observations are independent.

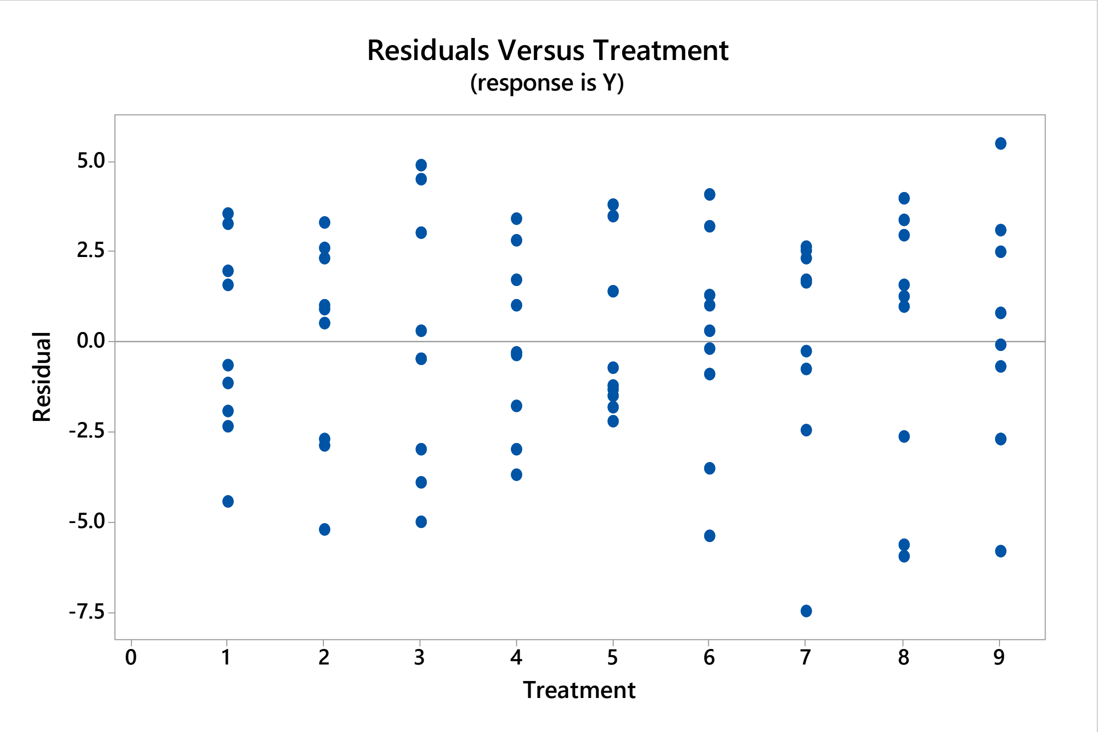


Test these assumptions with residuals diagnostic plots:

- Normal probability plot of the residuals.
- Plot of the residuals vs. treatments.
- Plot of the residuals vs. the predicted values.
- Plot of the residuals vs. the run order.

ANOVA Assumptions





ANOVA Sums of Squares

ANOVA separates the total variation in the data set into components attributed to different sources. The total amount of variation in the data set is:

$$SS_{total} = \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{\bar{y}})^2$$

If the k treatment means are $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$, that is:

$$\bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij}$$

then

$$\begin{aligned} SS_{total} &= \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j + \bar{y}_j - \bar{\bar{y}})^2 \\ &= \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2 + n \sum_{j=1}^{k} (\bar{y}_j - \bar{\bar{y}})^2 \\ &= SS_{\epsilon} + SS_{treatment} \end{aligned}$$

The degrees of freedom are also partitioned:

$$\begin{aligned} df_{total} &= df_{treatment} + df_{\epsilon} \\ kn - 1 &= (k-1) + k(n-1) \end{aligned}$$

The required variances, also called mean squares (MS), are given by:

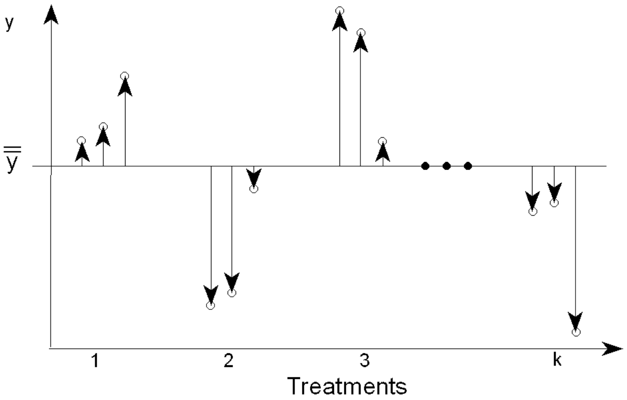
$$MS_{\epsilon} = s_{\epsilon}^2 = \frac{SS_{\epsilon}}{df_{\epsilon}} \text{ and } MS_{treatment} = n s_{\bar{y}}^2 = \frac{SS_{treatment}}{df_{treatment}}$$

so

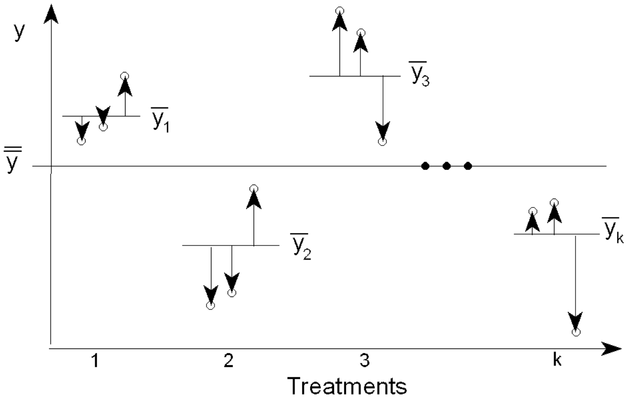
$$F = \frac{n s_{\bar{y}}^2}{s_{\epsilon}^2} = \frac{MS_{treatment}}{MS_{\epsilon}}$$

The statistic F follows an F distribution with $df_{numerator} = k - 1$ and $df_{denominator} = k(n-1)$. If $H_0 : \mu_i = \mu_j$ is true then $E(F) = 1$. If H_0 is false then $E(F) > 1$. We accept or reject H_0 on the basis of where F falls with respect to F_{α}.

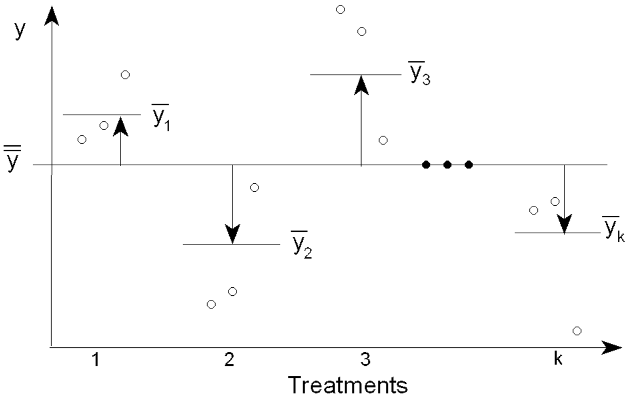
Total Variation: $SS_{total} = \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{\bar{y}})^2$



Error Variation: $SS_{\epsilon} = \sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$



Variation Between Treatments: $SS_{treatment} = n \sum_{j=1}^{k} (\bar{y}_j - \bar{\bar{y}})^2$



The ANOVA Table

Source	df	SS	MS	F
Treatment (A)	$k-1$	SS_A	SS_A/df_A	MS_A/MS_ϵ
Error	$k(n-1)$	SS_ϵ	SS_ϵ/df_ϵ	
Total	$kn-1$	SS_{total}		

ANOVA Summary Statistics

- Standard error of the model:

$$s_\epsilon = \sqrt{MS_\epsilon} = \sqrt{\frac{SS_\epsilon}{df_\epsilon}} = \sqrt{\frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij}-\bar{y}_j)^2}{k(n-1)}}$$

- Coefficient of determination:

$$r^2 = \frac{SS_{treatment}}{SS_{total}} = 1 - \frac{SS_\epsilon}{SS_{total}}$$

- Adjusted coefficient of determination:

$$r^2_{adj} = 1 - \left(\frac{df_{total}}{df_\epsilon}\right)\left(\frac{SS_\epsilon}{SS_{total}}\right)$$

Randomization

For an experiment to compare three processes (A, B, and C), what run order (1, 2, 3, or 4) should be used to collect the data?

Method	Run Order
1	AAAAAABBBBBBCCCCCC
2	AAABBBCCCAAABBBCCC
3	BBBAAABBBCCCAAACCC
4	CBCAABCCCABBAABCAB

- What if an unobserved lurking variable that affects the response changes during the experiment?

L	111112222233333333

- The ANOVA to test for differences between A, B, and C does not depend on or account for the run order ...
- However, the interpretation of the results does.
- Conclude that it is essential to randomize the run order.
- Method #4 is called the completely randomized design (CRD)
- If you do not randomize the run order your interpretation of the ANOVA may be incorrect and is open to challenge.

Post-ANOVA Pairwise Tests of Means

Although ANOVA indicates if there are significant differences between treatment means, it does not identify which pairs are different. Special pairwise testing methods are used after ANOVA:

- Two-sample t tests are too risky because of compounded testing errors
- 95% confidence intervals
- Bonferroni’s method - reduce α by the number of tests n, i.e. $\alpha' = \alpha/n$
- Sidak’s Method - less conservative than Bonferroni’s method
- Duncan’s Multiple Range Test - very sensitive, but a bit tedious
- Tukey’s Method (Tukey-Kramer or Tukey HSD) - popular
- Dunnett’s Method - for comparison to a control
- Hsu’s Method - for comparison against the best (highest or lowest) among the available treatments

One-Way ANOVA in MINITAB

- Use **Stat**> **ANOVA**> **One-way** if the response is in a single column (i.e. stacked) with an associated ID column.
- Use **Stat**> **ANOVA**> **One-way** (**Unstacked**) if each treatment is in its own column.
- In the **Graphs** menu:
  - Histogram and normal plot of the residuals.
  - Residuals vs. fits.
  - Residuals vs. order.
  - Residuals vs. the independent variable.
- In the **Comparisons** menu
  - Tukey’s method for all possible comparisons while controlling the family error rate.
  - Fisher’s method with a specified α (e.g. Bonferroni correction) for a specific subset of all possible tests.
  - Dunnett’s method for comparison against a control.
  - Hsu’s method for comparison against the best (highest or lowest) of the treatments.

One-way ANOVA in NCSS

Use **Analysis**> **ANOVA**> **One-way ANOVA**:

- On the **Variables** tab:
  - Set the **Response Variable**
  - Set the **Factor Variable**
- On the **Reports** tab turn on the:
  - **Assumptions Report**
  - **ANOVA Report**
  - **Means Report**
  - **Means Plot**
  - **Box Plots**
  - **Tukey-Kramer Test**

Response Transformations

If the ANOVA assumptions of homoscedasticity and/or normality of the residuals are not satisifed then it might be possible to transform the values of the response so that the assumptions are satisfied. In general, transformations take the form $y' = f(y)$ such as:

- $y' = \sqrt{y}$
- $y' = ln(y)$ or $y' = log(y)$
- $y' = y^2$
- $y' = y^\lambda$ where λ is chosen to make y' as normal as possible (Box-Cox transform)
- $y' = e^y$ or $y' = 10^y$
- For count data: $y' = \sqrt{y}$
- For proportions: $p' = \arcsin\left(\sqrt{p}\right)$
- If a suitable transform cannot be found but the residuals are non-normal but identically distributed (i.e. homoscedastic and same shape) then use the Kruskal-Wallis method by replacing the response with the ranked response, that is:

$$y' = rank(y)$$

Transformations in MINITAB

- Perform transformations from the **Calc**> **Calculator** menu or use the **let** command at the command prompt. For example:

```
mtb> let c3 = sqrt(c2)
```

Transformations in NCSS

- Enter the transformation in the **Transformation** column of the **Variable Info** tab, e.g. $sqrt(c1)$. Then select **Data**> **Recalc All** or click the calculator icon to apply the transformation.

Sample Size Calculation for One-way ANOVA

There is an exact calculation of the sample size for the ANOVA's F test presented in the text book; however, a simple and approximate sample size for a one-way classification design can be obtained by applying a Bonferroni correction to the type 1 error rate (α) for the two-sample t test.

- Recall from Chapter 3 that the sample size n per treatment group for the two-sample t test is given by:

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

  where δ is the effect size $(\delta = \mu_1 - \mu_2)$ with associated power/probability $P = 1 - \beta$ where β is the type 2 error rate and α is the type 1 error rate.
- In a one-way classification design with k treatments there will be $\binom{k}{2}$ multiple comparisons tests. By Bonferroni's method: To limit the family error rate to α_{family} the type 1 error rate for each test must be

$$\alpha' = \frac{\alpha_{family}}{\binom{k}{2}} = \frac{2\alpha_{family}}{k(k-1)}$$

  and the sample size per treatment group must be

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

Chapter 6: Experiments for Multi-way Classifications

Two Way Classification Problem

There are a levels of the first variable A (in columns) and b levels of the second B (in rows):

		A				
	$\mathbf{y}_{ij}$	**1**	**2**	**3**	$\cdots$	**a**
B	**1**	y_{11}	y_{21}	y_{31}	$\cdots$	y_{a1}
	2	y_{12}	y_{22}	y_{32}	$\cdots$	y_{a2}
	3	y_{13}	y_{23}	y_{33}	$\vdots$	y_{a3}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	b	y_{1b}	y_{2b}	y_{3b}	$\cdots$	y_{ab}

The model we will apply is:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where the α_i quantify the differences between the columns and the β_j quantify the differences between the rows.

Two-way ANOVA Hypotheses

The hypotheses to be tested are:

$$H_0 : \alpha_i = 0 \text{ for all of the } i$$

$$H_A : \alpha_i \neq 0 \text{ for at least one of the } i$$

$$H_0 : \beta_j = 0 \text{ for all of the } j$$

$$H_A : \beta_j \neq 0 \text{ for at least one of the } j$$

This will require two separate tests from the same two-way classified data set.

The Variable Effects

Analogous to the one-way ANOVA:

$$s_\alpha^2 = \frac{\sum_{i=1}^{a} \hat{\alpha}_i^2}{a-1}$$

and

$$s_\beta^2 = \frac{\sum_{j=1}^{b} \hat{\beta}_j^2}{b-1}$$

The error variance calculated from the ϵ_{ij} :

$$s_\epsilon^2 = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2}{(a-1)(b-1)}$$

where

$$\epsilon_{ij} = y_{ij} - (\mu + \alpha_i + \beta_j)$$

Tests for Variable Effects

By ANOVA:

$$F_A = \frac{bs_\alpha^2}{s_{error}^2}$$

with $(a-1)$ and $(a-1)(b-1)$ degrees of freedom for the numerator and denominator, respectively.

$$F_B = \frac{as_\beta^2}{s_{error}^2}$$

with $(b-1)$ and $(a-1)(b-1)$ degrees of freedom for the numerator and denominator, respectively.

Example

For the following two-way classification problem determine the row and column effects and use them to determine the row and column F ratios. Are they significant at $\alpha = 0.01$? There are four levels of the column variable A and three levels of the row variable B.

		A			
	$\mathbf{y}_{ij}$	**1**	**2**	**3**	**4**
B	**1**	18	42	34	46
	2	16	40	30	42
	3	11	35	29	41

Solution: The row and column means are:

		A				
	$\mathbf{y}_{ij}$	**1**	**2**	**3**	**4**	**Mean**
B	**1**	18	42	34	46	$\bar{y}_{\bullet 1} = 35$
	2	16	40	30	42	$\bar{y}_{\bullet 2} = 32$
	3	11	35	29	41	$\bar{y}_{\bullet 3} = 29$
	Mean	$\bar{y}_{1\bullet} = 15$	$\bar{y}_{2\bullet} = 39$	$\bar{y}_{3\bullet} = 31$	$\bar{y}_{4\bullet} = 43$	$\bar{\bar{y}} = 32$

The row and column effects, α_i and β_j, respectively, are the differences between the row and column means and the grand mean:

		A					
	$\mathbf{y}_{ij}$	**1**	**2**	**3**	**4**	**Mean**	$\hat{\beta}_j$
B	**1**	18	42	34	46	$\bar{y}_{\bullet 1} = 35$	$\hat{\beta}_1 = 3$
	2	16	40	30	42	$\bar{y}_{\bullet 2} = 32$	$\hat{\beta}_2 = 0$
	3	11	35	29	41	$\bar{y}_{\bullet 3} = 29$	$\hat{\beta}_3 = -3$
	Mean	$\bar{y}_{1\bullet} = 15$	$\bar{y}_{2\bullet} = 39$	$\bar{y}_{3\bullet} = 31$	$\bar{y}_{4\bullet} = 43$	$\bar{\bar{y}} = 32$	$\bar{\beta} = 0$
	$\hat{\alpha}_i$	$\hat{\alpha}_1 = -17$	$\hat{\alpha}_2 = 7$	$\hat{\alpha}_3 = -1$	$\hat{\alpha}_4 = 11$	$\bar{\alpha} = 0$	

Notice that the mean column and row effects are $\bar{\alpha} = 0$ and $\bar{\beta} = 0$ as required.

The effect variances are given by:

$$\begin{aligned} s_\alpha^2 &= \tfrac{1}{a-1} \textstyle\sum_{i=1}^{a} \hat{\alpha}_i^2 \\ &= \tfrac{1}{4-1} \left((-17)^2 + (7)^2 + (-1)^2 + (11)^2 \right) \\ &= 153.3 \end{aligned}$$

and

$$\begin{aligned} s_\beta^2 &= \tfrac{1}{b-1} \textstyle\sum_{j=1}^{b} \hat{\beta}_j^2 \\ &= \tfrac{1}{3-1} \left((3)^2 + (0)^2 + (-3)^2 \right) \\ &= 9.0 \end{aligned}$$

The matrix of errors is:

		A			
	ϵ_{ij}	**1**	**2**	**3**	**4**
B	**1**	0	0	0	0
	2	1	1	-1	-1
	3	-1	-1	1	1

Notice that the row and column sums add up to 0 as required.
The error variance is given by:

$$\begin{aligned} s_{error}^2 &= \tfrac{1}{(a-1)(b-1)} \textstyle\sum_{i=1}^{a} \sum_{j=1}^{b} \epsilon_{ij}^2 \\ &= \tfrac{1}{(4-1)(3-1)} \left((0)^2 + (0^2) + \cdots + (1)^2 \right) \\ &= 1.33 \end{aligned}$$

Finally the F ratio for the A effect is:

$$\begin{aligned} F_A &= \tfrac{b s_\alpha^2}{s_{error}^2} \\ &= \tfrac{3 \times 153.3}{1.33} \\ &= \tfrac{460}{1.33} \\ &= 346 \end{aligned}$$

and the F ratio for the B effect is:

$$\begin{aligned} F_B &= \tfrac{a s_\beta^2}{s_{error}^2} \\ &= \tfrac{4 \times 9.0}{1.33} \\ &= \tfrac{36}{1.33} \\ &= 27.1 \end{aligned}$$

The ANOVA Table (One Replicate)

Source	**df**	**SS**	**MS**	**F**
A	$a-1$	SS_A	MS_A	MS_A/MS_ϵ
B	$b-1$	SS_B	MS_B	MS_B/MS_ϵ
Error	$(a-1)(b-1)$	SS_ϵ	MS_ϵ	
Total	$ab-1$	SS_{total}		

Multi-way ANOVA in MINITAB

- Use **Stat**> **ANOVA**> **Two-Way** for two-way classifications.
- Use **Stat**> **ANOVA**> **Balanced ANOVA** for balanced multi-way classifications.
- Use **Stat**> **ANOVA**> **General Linear Model** for almost everything.
  - Select residuals diagnostic graphs from the **Graphs** menu.
  - Select an appropriate post-ANOVA comparisons method from the **Comparisons** menu.
  - Be careful how you interpret the F statistics!

Multi-way ANOVA in NCSS

Analysis> **ANOVA**> **Analysis of Variance**

- On the **Variables** Tab:
  - Set the **Response Variable**
  - Set the **Factor 1**, **2**, ..., **Variables**
- On the **Reports** Tab:
  - **ANOVA Report**
  - **Means Report**
  - **Means Plots**
  - **Tukey-Kramer Test**

Blocking

Suppose that we want to test three different processes A, B, and C for possible differences between their means but we know there is lots of noise so we will have to take several observations from each process. Which of the following run orders should be used to collect the data?

Method	Run Order
1	AAAAAABBBBBBCCCCCC
2	AAABBBCCCAAABBBCCC
3	BBBAAABBBCCCAAACCC
4	CBCAABCCCABBAABCAB

What if the process is unstable and drifts significantly over the time period required to collect the data? If this drift is not handled correctly it may hide significant differences between the three processes or its effect might be misattributed to differences between the three processes.

The solution is to build the experiment in blocks which can be used to remove the effect of the drift. Such designs are called *randomized block designs* (RBD).

Method	Run Order (Blocked)
5	ABACCACBB | CBAAACBBC
6	BCCAAB | CABABC | ABCACB
7	BCA | ACB | CAB | BAC | CBA | ABC

The two-way ANOVA will test for differences between A, B, and C while controlling for differences between blocks so conditions should be homogeneous within blocks but may be heterogeneous between blocks. There are many opportunities to improve experiments with the use of blocking to control unavoidable sources of variation.

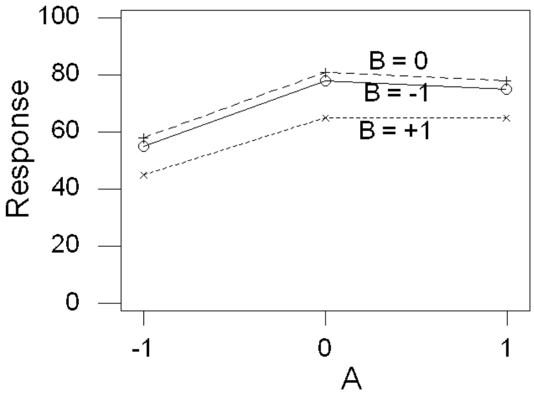
The following table shows how the degrees of freedom will be allocated in the various models:

	Method			
	4	5	6	7
Block	0	1	2	5
Treatment	2	2	2	2
Error	15	14	13	10
Total	17	17	17	17

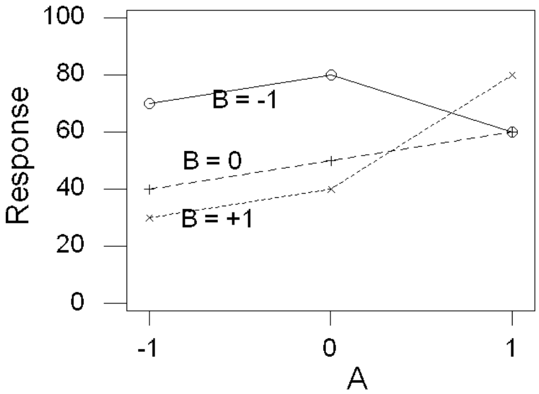
Interactions

When two variables interact then the effect of one variable depends on the level of the other. In case a) below A and B do not interact. In case b) below A and B do interact. In general, in such plots (often called interaction plots), parallel line segments over all vertical slices in the plot indicate no interaction and divergent line segments over some or all vertical slices in the plot indicate interaction.

a.



b.



To be capable of detecting an interaction a two-way factorial experiment requires two or more replicates of the $a \times b$ design.

The ANOVA Table with Interaction

In an $a \times b$ factorial experiment with n replicates:

Source	df	SS	MS	F
A	$a - 1$	SS_A	MS_A	MS_A/MS_ϵ
B	$b - 1$	SS_B	MS_B	MS_B/MS_ϵ
AB	$(a-1)(b-1)$	SS_{AB}	MS_{AB}	MS_{AB}/MS_ϵ
Error	$ab(n-1)$	SS_ϵ	MS_ϵ	
Total	$nab - 1$	SS_{total}		

Higher Order Interactions

When there are more than two variables then three-factor, four-factor, and higher order interactions are possible. In most engineering technologies three-factor and higher order interactions are rare and it is safe to ignore them. In some technologies (like psychology) high order interactions can be very important.

ANOVA for the Three-way Classification Design

In an $a \times b \times c$ factorial experiment with n replicates:

Source	df	SS	MS	F
A	$a-1$	SS_A	MS_A	MS_A/MS_ϵ
B	$b-1$	SS_B	MS_B	MS_B/MS_ϵ
C	$c-1$	SS_C	MS_C	MS_C/MS_ϵ
AB	$(a-1)(b-1)$	SS_{AB}	MS_{AB}	MS_{AB}/MS_ϵ
AC	$(a-1)(c-1)$	SS_{AC}	MS_{AC}	MS_{AC}/MS_ϵ
BC	$(b-1)(c-1)$	SS_{BC}	MS_{BC}	MS_{BC}/MS_ϵ
ABC	$(a-1)(b-1)(c-1)$	SS_{ABC}	MS_{ABC}	MS_{ABC}/MS_ϵ
Error	$abc(n-1)$	SS_ϵ	MS_ϵ	
Total	$nabc-1$	SS_{total}		

The df and SS associated with any insignificant terms that are omitted or dropped from the model are pooled with df_ϵ and SS_ϵ, respectively. When insignificant terms are dropped from the model, they must be managed to preserve the hierarchy of the remaining terms in the model. For example, in order to retain the BCE three-factor interaction in the model it's necessary to retain B, C, E, BC, BE, and CE even if they are not all statistically significant.

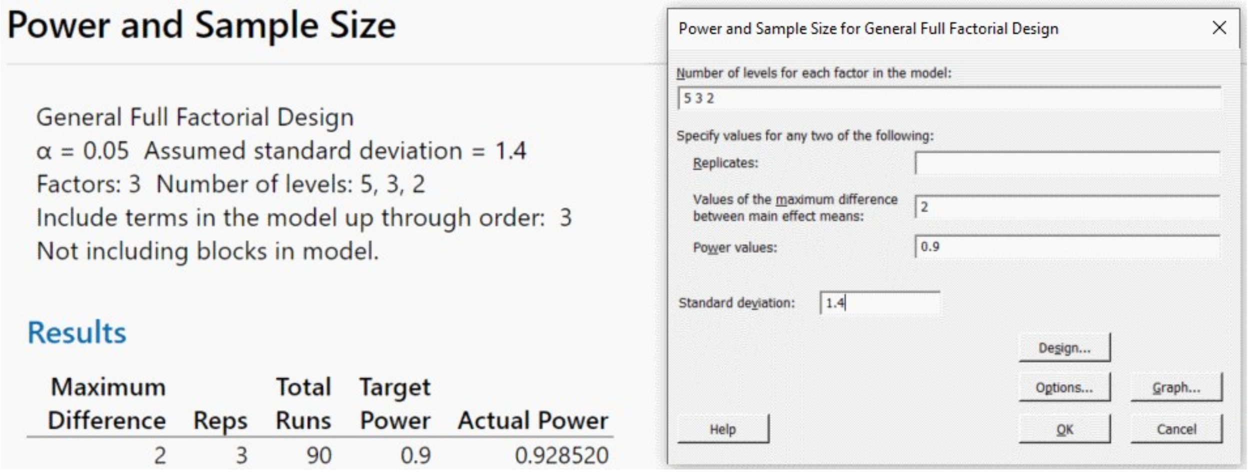
Sample Size Calculations

- In a two-way or multi-way classification design, if the experiment must be able to resolve a specified effect size with specified power between pairs of levels for all of the study variables, then the variable with the largest number of levels will be the limiting case because it will have the fewest observations in each of its levels. The power for the other variables with fewer levels will be greater than the specified power because they will have more observations per level.
- Sample size calculations for two-way and multi-way classification designs:
  - Are closely related in method and result to the sample size calculations for one-way classification designs and two-sample t tests so can be approximated by those methods.
  - Can be performed exactly for ANOVA F tests using MINITAB **Stat**> **Power and Sample Size**> **General Full Factorial Design**.

Sample Size Calculations

Example: Determine the number of replicates required for a $5 \times 3 \times 2$ full factorial experiment if the experiment must be capable of detecting an effect of size $\delta = 2$ with 90% power. The standard error is expected to be $\sigma_\epsilon = 1.4$.

Solution 1: Using **Stat**> **Power and Sample Size**> **General Full Factorial Design** the experiment will require three replicates and the power to detect the effect of size $\delta = 2$ will be 92.8% for the five-level variable. The total number of runs required for the experiment will be $5 \times 3 \times 2 \times 3 = 90$.



```
Power and Sample Size

General Full Factorial Design
α = 0.05  Assumed standard deviation = 1.4
Factors: 3  Number of levels: 5, 3, 2
Include terms in the model up through order:  3
Not including blocks in model.

Results

  Maximum          Total  Target
  Difference  Reps  Runs   Power  Actual Power
           2     3    90     0.9      0.928520
```

Power and Sample Size for General Full Factorial Design

Number of levels for each factor in the model: 5 3 2

Specify values for any two of the following:

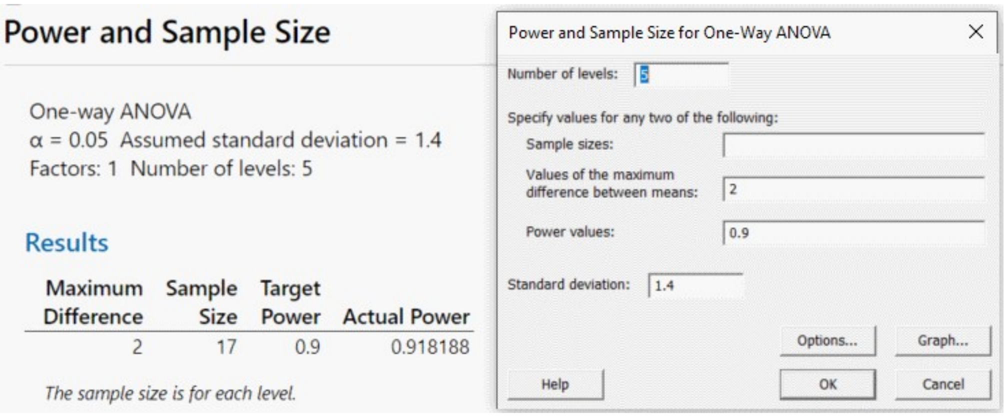
Replicates:

Values of the maximum difference between main effect means: 2

Power values: 0.9

Standard deviation: 1.4

Solution 2: Using **Stat**> **Power and Sample Size**> **One-way ANOVA** for the five level variable the experiment will require $5 \times 17 = 85$ runs - in good agreement with the 90 runs calculated in the first solution.



```
Power and Sample Size

One-way ANOVA
α = 0.05  Assumed standard deviation = 1.4
Factors: 1  Number of levels: 5

Results

  Maximum     Sample  Target
  Difference    Size   Power  Actual Power
           2      17     0.9      0.918188

  The sample size is for each level.
```

Power and Sample Size for One-Way ANOVA

Number of levels: 5

Specify values for any two of the following:

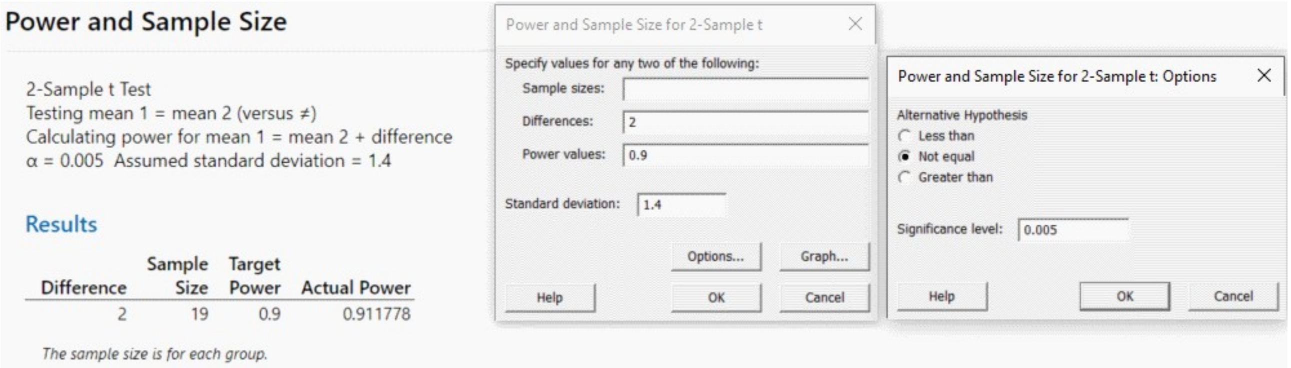
Sample sizes:

Values of the maximum difference between means: 2

Power values: 0.9

Standard deviation: 1.4

Solution 3: Using **Stat**> **Power and Sample Size**> **Two-sample t** applied to the five-level variable with a Bonferroni correction for $\binom{5}{2} = 10$ tests (i.e. $\alpha' = 0.05/10 = 0.005$) gives an experiment with 19 observations per group or $5 \times 19 = 95$ total observations. The sample size is a bit larger than the others because the Bonferroni correction is conservative.



```
Power and Sample Size

2-Sample t Test
Testing mean 1 = mean 2 (versus ≠)
Calculating power for mean 1 = mean 2 + difference
α = 0.005  Assumed standard deviation = 1.4

Results

              Sample  Target
  Difference    Size   Power  Actual Power
           2      19     0.9      0.911778

  The sample size is for each group.
```

Power and Sample Size for 2-Sample t

Specify values for any two of the following:

Sample sizes:

Differences: 2

Power values: 0.9

Standard deviation: 1.4

Power and Sample Size for 2-Sample t: Options

Alternative Hypothesis

○ Less than

● Not equal

○ Greater than

Significance level: 0.005

Chapter 7: Advanced ANOVA Topics

Balanced Incomplete Factorial Designs

- Full-factorial designs include all possible permutations of all levels of the design variables.
- Full-factorial designs can resolve main effects, two-factor interactions, and higher order interactions.
- Balanced incomplete factorial designs omit some of the runs from the full-factorial design to decrease the number of runs required for the experiment.
- The runs are omitted uniformly to preserve the balance of the experiment, i.e. all levels of each variable are equally represented.
- Balanced incomplete factorial designs can only resolve main effects and their accuracy depends on the assumption that there are no significant two-factor and higher order interactions.

Example: Consider the 3×3 balanced incomplete factorial design:

		A		
		1	2	3
B	1	✓	✓	×
	2	✓	×	✓
	3	×	✓	✓

Latin Squares

- Latin squares are balanced incomplete designs with three variables.
- All variables have the same number of levels $n = 3, 4, \ldots$ but only $1/n$ of the possible runs from the full-factorial design are used.
- Can only resolve main effects and assume (rightly or not) that there are no significant interactions.
- Usually employed as a blocking design to study one variable (C) and block two others $(A$ and $B)$.

Example: Consider the 3×3 Latin Square design:

		B		
		$\mathbf{B}_1$	$\mathbf{B}_2$	$\mathbf{B}_3$
A	$\mathbf{A}_1$	C_2	C_3	C_1
	$\mathbf{A}_2$	C_3	C_1	C_2
	$\mathbf{A}_3$	C_1	C_2	C_3

Fixed and Random Variables

- Suppose that we need to do an experiment to study biases between operators for a critical measurement.
  - If there only a few operators who are qualified to make the measurement and all of the operators can be included in the study, then it makes sense to test for biases between operators:

    $$H_0 : \mu_i = \mu_j \text{ for all possible } i,j$$

    $$H_A : \mu_i \neq \mu_j \text{ for at least one } i,j \text{ pair}$$

    If there are significant biases between operators, then future measurements could be corrected by removing the known biases. In this case the qualitative variable Operator is called a *fixed* variable, factor, or effect.
  - If there are many operators - too many to include all of them in the study - then only a random sample of the operators can be included in the study. In this case it makes sense to test the standard deviation of biases between operators:

    $$H_0 : \sigma_{Operator} = 0$$

    $$H_A : \sigma_{Operator} > 0$$

    If the standard deviation of biases between operators is non-zero and practically large, then the measurements will be contaminated by noise from the operators and it may be necessary to take action to reduce operator biases, such as by training the operators or by replacing the instrument with one that is less sensitive to how it is used. In this case the qualitative variable Operator is called a *random* variable, factor, or effect.
- The analysis of a qualitative variable as fixed or random affects the ANOVA, i.e. the ANOVA results obtained by treating a variable as fixed or random are different, so it is essential that the fixed or random nature of a variable be specified correctly before running the ANOVA.
  - The choice to treat a variable as fixed or random comes down to the action taken if the effect is statistically and practically significant:
    - If you can act on the individual differences then the variable is fixed, e.g. "recalibrating" a single operator
    - If you have to act on the whole population then the variable is random, e.g. retraining all of the operators
- Fixed and Random Variables in MINITAB:
  - In MINITAB’s **Stat**> **ANOVA**> **General Linear Model**> **Fit General Linear Model** menu MINITAB assumes that a qualitative variable is a fixed factor. To change a variable’s fixed status to random open the **Random/Nest** submenu and in the **Factor type**: table choose the appropriate setting, **Fixed** or **Random**, from the list box.
  - By default MINITAB doesn’t report the standard deviations of the random effects. To obtain that report, open the **Results** submenu and check the checkbox next to **Variance components**.

Analysis of Fixed and Random Variables

- If A is fixed and B is fixed:

Source	df	$E(MS)$	F
A	$a-1$	$\sigma_\epsilon^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2$	$\frac{MS_A}{MS_\epsilon}$
B	$b-1$	$\sigma_\epsilon^2 + \frac{an}{b-1}\sum_{j=1}^{b}\beta_j^2$	$\frac{MS_B}{MS_\epsilon}$
AB	$(a-1)(b-1)$	$\sigma_\epsilon^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}^2$	$\frac{MS_{AB}}{MS_\epsilon}$
$Error(\epsilon)$	$ab(n-1)$	σ_ϵ^2	
$Total$	$abn-1$		

Analysis of Fixed and Random Variables

- If A is fixed and B is random:

Source	df	$E(MS)$	F
A	$a-1$	$\sigma_\epsilon^2 + n\sigma_{AB}^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2$	$\frac{MS_A}{MS_{AB}}$
B	$b-1$	$\sigma_\epsilon^2 + n\sigma_{AB}^2 + an\sigma_B^2$	$\frac{MS_B}{MS_{AB}}$
AB	$(a-1)(b-1)$	$\sigma_\epsilon^2 + n\sigma_{AB}^2$	$\frac{MS_{AB}}{MS_\epsilon}$
$Error(\epsilon)$	$ab(n-1)$	σ_ϵ^2	
$Total$	$abn-1$		

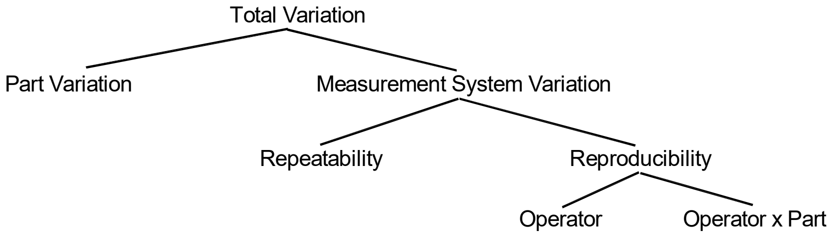
Analysis of Fixed and Random Variables

- If A is random and B is random:

$Source$	df	$E(MS)$	F
A	$a-1$	$\sigma_\epsilon^2 + n\sigma_{AB}^2 + bn\sigma_A^2$	$\frac{MS_A}{MS_{AB}}$
B	$b-1$	$\sigma_\epsilon^2 + n\sigma_{AB}^2 + an\sigma_B^2$	$\frac{MS_B}{MS_{AB}}$
AB	$(a-1)(b-1)$	$\sigma_\epsilon^2 + n\sigma_{AB}^2$	$\frac{MS_{AB}}{MS_\epsilon}$
$Error(\epsilon)$	$ab(n-1)$	σ_ϵ^2	
$Total$	$abn-1$		

Gage Error Studies

- Measurement *accuracy* is established by calibration.
- Measurement *precision* is quantified in a designed experiment called a *gage error study* (GR&R study). The purpose of the GR&R study is to obtain estimates of the different sources of variability in the measurement system:



- In a typical gage error study three or more operators measure the same ten parts two times.
- If the operators are fixed and if a difference between operators is detected we might adjust the present and future data for operator bias or ‘calibrate’ one or more of the operators.
- If the operators are random and if σ_{Op} is determined to be too large we would have to train all of the operators, not just those who participated in the study. It would be inappropriate to take any action against specific operators who participated in the study.
- In most gage error studies operators are assumed to be a random sample from many possible operators. Then ANOVA can be used to partition the total observed variability in the gage error study data into three components: part variation, operator variation (reproducibility), and inherent measurement error (repeatability or precision):

Source	*df*	*MS*	*E*(*MS*)	*F*
Operator(*O*)	$o-1$	MS_O	$\sigma_\epsilon^2 + np\sigma_O^2$	$\frac{MS_O}{MS_\epsilon}$
Part(*P*)	$p-1$	MS_P	$\sigma_\epsilon^2 + no\sigma_P^2$	$\frac{MS_P}{MS_\epsilon}$
Error(ϵ)	$opn-o-p+1$	MS_ϵ	σ_ϵ^2	
Total	$opn-1$			

These variances are determined using a post-ANOVA method called *variance components analysis*:

$$\hat{\sigma}_\epsilon^2 = MS_\epsilon$$

$$\hat{\sigma}_{OP}^2 = \frac{MS_{OP} - MS_\epsilon}{n}$$

$$\hat{\sigma}_O^2 = \frac{MS_O - MS_{OP}}{np}$$

$$\hat{\sigma}_P^2 = \frac{MS_P - MS_{OP}}{no}$$

- After the σs are known from the variance components analysis they are used to calculate quantities called the equipment variation (EV) which estimates precision and the appraiser variation (AV) which estimates reproducibility from:

  $$EV = 6\sigma_{\epsilon}$$

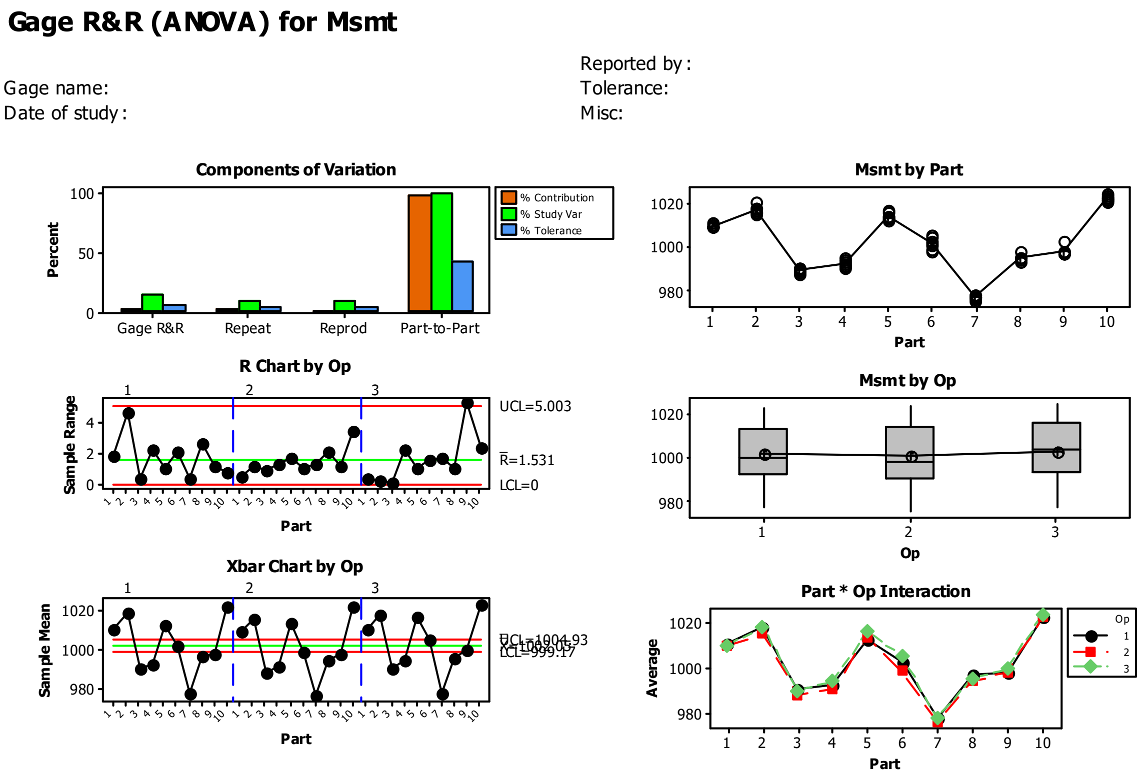
  $$AV = 6\sigma_{Op}$$

  The 6σ value comes from the normal distribution - about 99.7% of a normal distribution should fall within $\pm 3\sigma$ of the population mean which is an interval with width 6σ wide.
- If both reproducibility (AV) and repeatability (EV) are less than about 10% of the tolerance then the measurement system, consisting of the operators, instrument, and measurement methods, is acceptable; if they are between 10% and 30% of the tolerance the measurement system is marginal; and if they are greater than 30% the measurement system should definitely not be used.

Sample Size in GR&R Studies

- Most GR&R study designs provide plenty of degrees of freedom for estimating repeatability but few to estimate operator reproducibility.
- Use enough parts to challenge the operators.
- A minimum of 6-8 operators is recommended. (See Burdick, Borror, and Montgomery, Design and Analysis of Gauge R&R Studies.)
- Each operator should measure each part twice. Three or more such trials only improve the repeatability estimate which is already precise compared to the reproducibility estimate.

GR&R Study Example



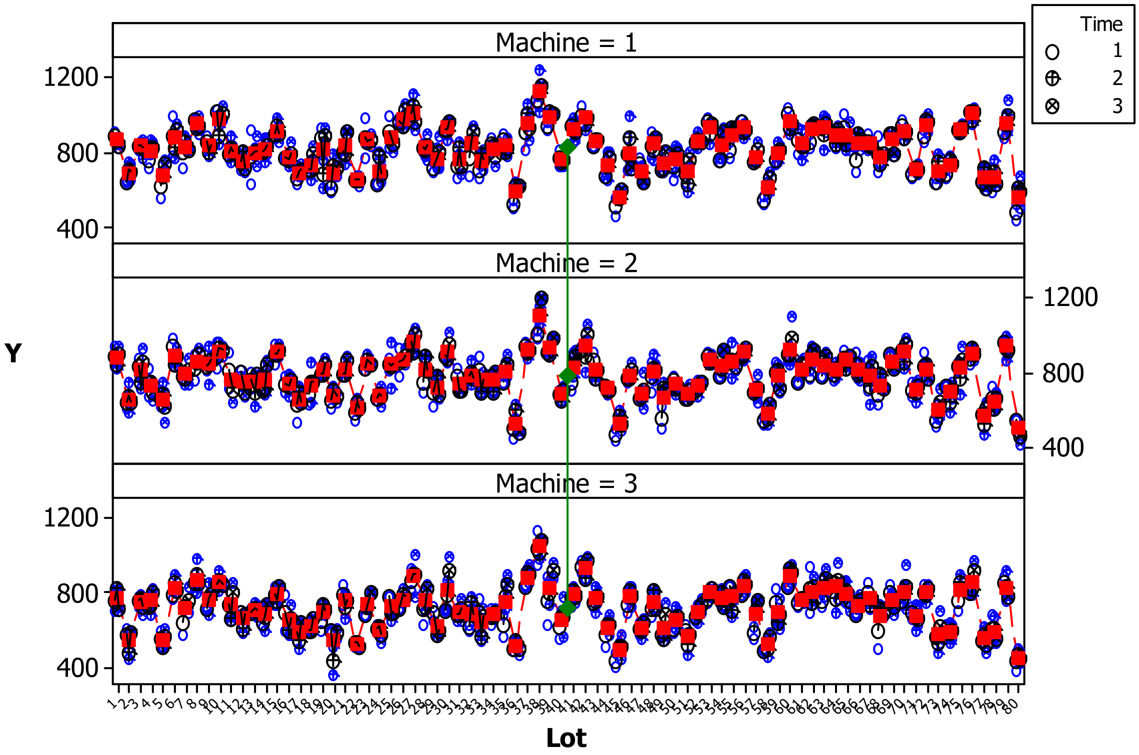
Gage R&R (ANOVA) for Msmt

Reported by:
Gage name:
Tolerance:
Date of study:
Misc:

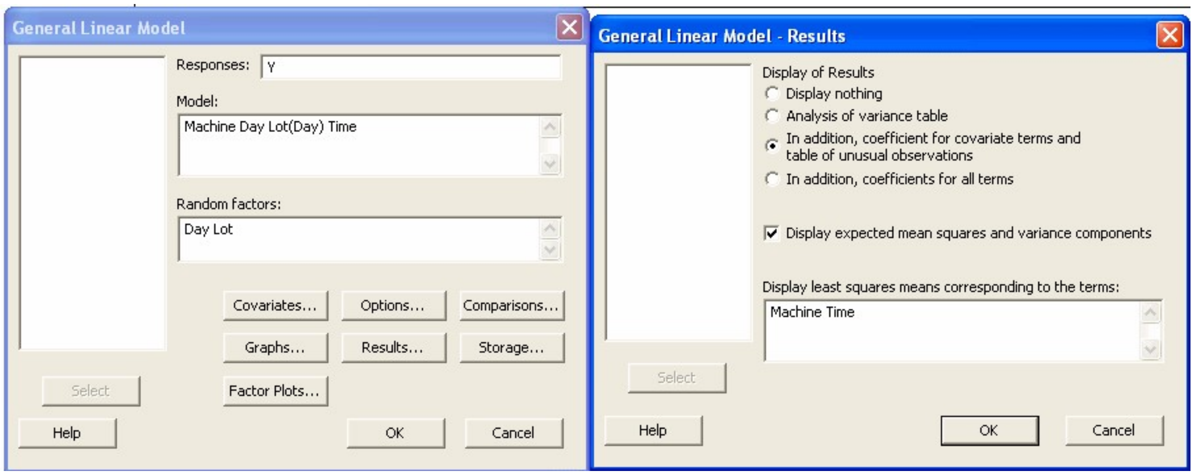
```
Gage R&R Study - ANOVA Method

Source          DF       SS       MS        F      P
Part             9  10700.0  1188.89  443.976  0.000
Op               2     55.8    27.92   10.427  0.001
Part * Op       18     48.2     2.68    1.442  0.183
Repeatability   30     55.7     1.86
Total           59  10859.8


                              %Contribution
Source              VarComp   (of VarComp)
Total Gage R&R        3.530           1.75
  Repeatability       1.858           0.92
  Reproducibility     1.672           0.83
    Op                1.262           0.63
    Op*Part           0.410           0.20
Part-To-Part        197.702          98.25
Total Variation     201.231         100.00


Process tolerance = 200


                                 Study Var  %Study Var  %Tolerance
Source             StdDev (SD)    (6 * SD)       (%SV)  (SV/Toler)
Total Gage R&R          1.8788     11.2728       13.24        5.64
  Repeatability         1.3629      8.1777        9.61        4.09
  Reproducibility       1.2932      7.7590        9.12        3.88
    Op                  1.1235      6.7408        7.92        3.37
    Op*Part             0.6404      3.8423        4.51        1.92
Part-To-Part           14.0606     84.3638       99.12       42.18
Total Variation        14.1856     85.1136      100.00       42.56


Number of Distinct Categories = 10
```

Variance Components in Process Capability Studies

Each lot of incoming material is split into three parallel paths to be processed on three hopefully identical machines. Four lots are processed each day for 40 days. The response is measured three times for each lot, once at the beginning, middle, and end. Two samples are measured at each time point.



Variance Components in Process Capability Studies



Responses: Y

Model:
Machine Day Lot(Day) Time

Random factors:
Day Lot

Display of Results
○ Display nothing
○ Analysis of variance table
● In addition, coefficient for covariate terms and table of unusual observations
○ In addition, coefficients for all terms

☑ Display expected mean squares and variance components

Display least squares means corresponding to the terms:
Machine Time

Variance Components in Process Capability Studies

```
General Linear Model: Y versus Machine, Day, Time, Lot

Factor    Type    Levels  Values
Machine   fixed        3  1, 2, 3
Day       random      20  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
                          16, 17, 18, 19, 20
Lot(Day)  random      80  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
                          16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,
                          28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,
                          40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51,
                          52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63,
                          64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75,
                          76, 77, 78, 79, 80
Time      fixed        3  1, 2, 3


Analysis of Variance for Y, using Adjusted SS for Tests

Source      DF    Seq SS    Adj SS   Adj MS       F      P
Machine      2   2612016   2612016  1306008  364.48  0.000
Day         19   6026897   6026897   317205    1.65  0.072
Lot(Day)    60  11507119  11507119   191785   53.52  0.000
Time         2    184803    184803    92401   25.79  0.000
Error     1356   4858866   4858866     3583
Total     1439  25189701

S = 59.8601   R-Sq = 80.71%   R-Sq(adj) = 79.53%
```

```
Variance Components, using Adjusted SS

          Estimated    Standard
Source     Variance   Deviation
Day            1742       41.74
Lot(Day)      10456      102.26
Error          3583       59.86


Least Squares Means for Y

Machine   Mean     Bias
1        826.4     47.7
2        786.7      8.0
3        723.0    -55.7
All      778.7      0.0

Time      Mean     Bias
1        766.3    -12.4
2        776.2     -2.5
3        793.7     15.0
All      778.7      0.0
```

Sample Size for Process Capability

An approximate $(1-\alpha)100\%$ confidence interval for c_p is given by

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

where the confidence interval's relative half-width is

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

Then the sample size required to obtain relative confidence interval half-width δ is

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

Example: The sample size required to estimate c_p with 10% precision and 95% confidence is

$$n = \frac{1}{2}\left(\frac{1.96}{0.1}\right)^2 = 192$$

Analysis of Experiments with Fixed and Random Variables in Minitab

Use **Stat**> **ANOVA**> **General Linear Model**. Enter all variables and terms in the **Model** window. Indicate the random variables in the **Random** window and continuous quantitative predictors as **Covariates**. Turn on **Display expected mean squares and variance components** in the **Results** window. Manually calculate the standard deviations from the variances in the MINITAB output.

Analysis of GR&R Studies in MINITAB

- MINITAB assumes that operators and parts are random per QS9000: Measurement Systems Analysis.
- Use **Stat**> **Quality Tools**> **Gage Study**> **Gage R&R Study** (**Crossed**) if all of the operators measure all of the parts.
- Use **Stat**> **Quality Tools**> **Gage Study**> **Gage R&R Study** (**Nested**) if each operator measures only his own parts.
- Specify the part's tolerance width in the **Options**> **Process Tolerance** window and MINITAB will report the usual relative variations.
- Complex GR&R studies that are structured according to the default crossed and nested designs should be analyzed using **Stat**> **ANOVA**< **General Linear Model**.

Analysis of Experiments with Fixed and Random Variables in NCSS

Use **Analysis**> **ANOVA**> **Analysis of Variance** or **Analysis**> **ANOVA**> **ANOVA GLM**. Set each variable's attribute, fixed or random, as required. NCSS performs the appropriate ANOVA and reports the variance components equations but does not solve them. You will have to solve them manually.

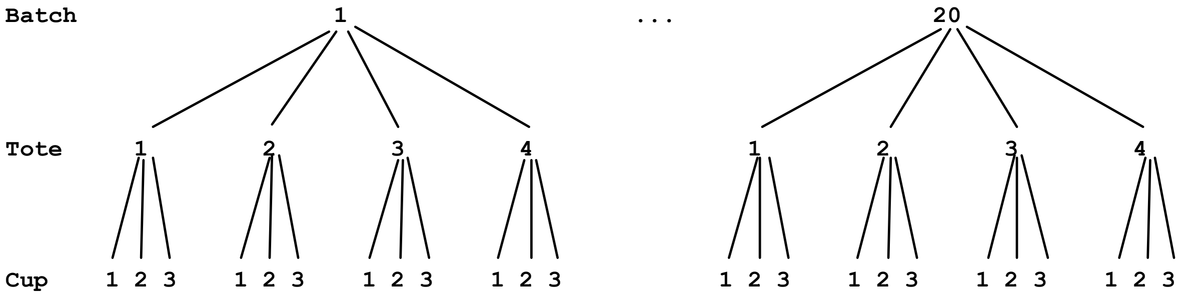
Analysis of GR&R Studies in NCSS

Assuming that operators and parts are both random and crossed (i.e. not nested) and each operator measures each part at least twice use **Analysis**> **Quality Control**> **R&R Study**. Given the part specifications NCSS will make the relevant comparisons between repeatability and reproducibility to the spec.

Nested Variables

Some experiments involve variables that have levels that are unique within the levels of other variables. The relationship between such variables is referred to as *nesting*.

Example: A dry powdered pharmaceutical product (active ingredient plus filler) is made in batches in an industrial blender. Each batch is unloaded into four totes and then material is vacuum-transfered into cups for packaging and distribution. An experiment was performed to study how much variability in the active ingredient comes from differences between batches, totes, and cups. The experiment included twenty batches, four totes per batch, and three cups were chosen at random from each tote and assayed for the active ingredient. A schematic and the analysis of the fully nested experiment design are shown below.



Data Display

```
Row  Batch  Tote  Cup      Msmt
  1      1     1    1  1063.50
  2      1     1    2  1062.87
  3      1     1    3  1059.63
  4      1     2    1  1054.66
  5      1     2    2  1054.03
  6      1     2    3  1050.79
.
.
.
234     20     2    3  1027.99
235     20     3    1  1066.86
236     20     3    2  1066.23
237     20     3    3  1062.99
238     20     4    1  1005.26
239     20     4    2  1004.63
240     20     4    3  1001.39
```

Nested ANOVA: Msmt versus Batch, Tote, Cup

```
Analysis of Variance for Msmt

Source   DF           SS         MS        F      P
Batch    19  146434.9677  7707.1036    3.980  0.000
Tote     60  116194.0506  1936.5675  448.539  0.000
Cup     160     690.8006     4.3175
Total   239  263319.8189


Variance Components

                        % of
Source  Var Comp.  Total   StDev
Batch     480.878  42.58  21.929
Tote      644.083  57.03  25.379
Cup         4.318   0.38   2.078
Total    1129.279         33.605
```

Analysis of Experiments With Nested Variables

Analyze fully nested designs in MINITAB using **Stat**> **ANOVA**> **Fully Nested Design** or **Stat**> **ANOVA**> **General Linear Model**. For the latter method, the example's model is specified as: `Batch Tote(Batch) Cup(Batch Tote)` although the last term should be dropped to provide error degrees of freedom for the analysis unless more than one assay is performed from each cup. The **Stat**> **ANOVA**> **General Linear Model** method can also be used to analyze complex designs with both crossed and nested variables.

Split-Plot Designs

- Split-plot designs are hybrid designs that cross a matrix of hard-to-change (HTC) variables with a matrix of easy-to-change variables (ETC) by nesting a design of the ETC variables within the runs of a design of the HTC variables.
- Split-plots apply different plans of randomization, blocking, repetitions, and replicates to the HTC and ETC variables.
- The levels of the hard-to-change variables are held constant within whole-plots, i.e. there is a *restriction on randomization*.
- The levels of the easy-to-change variables that define the split-plots are performed using complete randomization within each whole-plot; that is, split-plots are nested within whole-plots.
- The whole-plot to split-plot relationship is closely related to blocking in factorial design and repeated measures designs.
- Whole-plots and split-plots have different, independent randomization, blocking, and replication plans.
- In the ANOVA for a split-plot design, the whole-plots and split-plots have different estimates for the errors for calculating their F statistics. Consequently, ...
- The number of replicates for whole-plots is different from the number of replicates for split-plots.
- Warning: Many industrial experiments that were conceived as completely randomized factorial designs are executed as split-plot designs because of the presence of and complications associated with changing the hard-to-change variable levels. The analysis of an experiment executed as a split-plot but analyzed as a completely randomized factorial design will give incorrect results.

Example: A split-plot experiment will be performed with one HTC variable and one ETC variable. The HTC variable (A) has two levels and will use an RBD design with four replicates for eight whole-plot runs. The whole-plot run matrix is shown below.

```
Whole Plot Run Matrix

Block(A)  WP  A(HTC)
       1   2       2
       1   1       1
       2   3       1
       2   4       2
       3   5       1
       3   6       2
       4   7       1
       4   8       2
```

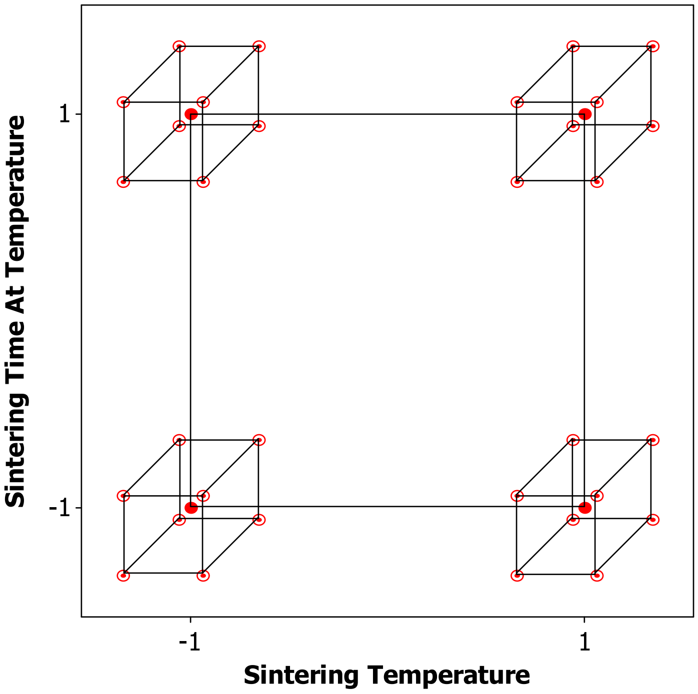
The ETC variable (B) has three levels of each variable and will use an RBD design with two replicates for six split-plot runs within each whole-plot. The split-plot run matrix is shown below in standard order. The complete experiment will have $8 \times 6 = 48$ runs.

```
Split-Plot Run Matrix (Standard Order)

Block(B)  B(ETC)
       1       1
       1       2
       1       3
       2       1
       2       2
       2       3
```

Split-Plot Designs

Example: An experiment will be performed to study the shrinkage (size reduction) of sintered ceramic parts as a function of:

- Hard-to-change / whole-plot variables (levels): Sintering temperature (2), Sintering time at temperature (2)
- Easy-to change / split-plot variables (levels): Ceramic grain size (2), binder amount (2), mold pressure (2)
- The experiment will have two replicates, built in blocks, of the 2^2 whole-plot design and four replicates, built in blocks, of the 2^3 split-plot design within each whole-plot for a total of $(2 \times 2^2) \times (4 \times 2^3) = 256$ runs. A schematic of one replicate of the whole-plot design and one replicate of the split-plot design is shown below.
- Each whole-plot, consisting of one of the split-plot cubes at one of the sintering temperature (A) by sintering time at temperature (B) combinations, will be completed before the next whole plot is started. Per the blocking on replicates requirement, the four whole-plots within one replicate of the 2^2 whole-plot design will be completed in random order before starting the second replicate of whole-plots.



The table below shows the randomization and blocking plan for the whole plots.

```
           WP
RO     Block   WP     A    B
 1         1    2     1    1
 2         1    1     1   -1
 3         1    3    -1   -1
 4         1    4    -1    1

 5         2    7    -1   -1
 6         2    8     1   -1
 7         2    5     1    1
 8         2    6    -1    1
```

Analysis of Split-Plot Designs

- In MINITAB use **Stat**> **DOE**> **Factorial**> **Create Factorial Design**> **2-level split-plot** to create a new split-plot design. Build the experiment and then use > **Analyze Factorial Design** to run the analysis.
- To analyze split-plot designs in MINITAB that are outside of its scope, use **Stat**> **ANOVA**> **General Linear Model** to perform the analysis. Use a column in the MINITAB worksheet to identify the whole-plots. Specify the whole-plot column as a random variable in the model. That column is necessary to build the error term for testing for whole-plot variable effects.

Example (from Poctner and Kowalski, *How To Analyze A Split-Plot Experiment*, Quality Progress, December 2004, p. 67-74.)

An experiment was performed to study the water resistance of stained wood as a function of pre-stain (a hard-to-change variable) and stain (an easy-to-change variable). There were two pre-stains and four stains. Pre-stains were applied to whole 4x8 foot sheets of plywood (the whole plots). Then each sheet of plywood was cut up into four pieces and each piece was painted with one of the stains (the split plots). The whole-plot design is 2^1 which was replicated three times (6 sheets of plywood). The split-plot design is 4^1 which was replicated one time within each whole-plot. The experimental runs and responses are shown in the table below. The P column indicates pre-stain, the S column indicates stain, and the WP column identifies the whole-plots. The analysis of the experiment is also shown in the table. To build the correct error terms for testing for whole-plot variable and split-plot variable effects, the model was specified as: **`P WP(P) S P*S and WP`** must be declared a random variable.

```
Row  P  S  WP     Y
  1  2  2   4  53.5
  2  2  4   4  32.5
  3  2  1   4  46.6
  4  2  3   4  35.4
  5  2  4   5  44.6
  6  2  1   5  52.2
  7  2  3   5  45.9
  8  2  2   5  48.3
  9  1  3   1  40.8
 10  1  1   1  43.0
 11  1  2   1  51.8
 12  1  4   1  45.5
 13  1  2   2  60.9
 14  1  4   2  55.3
 15  1  3   2  51.1
 16  1  1   2  57.4
 17  2  1   6  32.1
 18  2  4   6  30.1
 19  2  2   6  34.4
 20  2  3   6  32.2
 21  1  1   3  52.8
 22  1  3   3  51.7
 23  1  4   3  55.3
 24  1  2   3  59.2
```

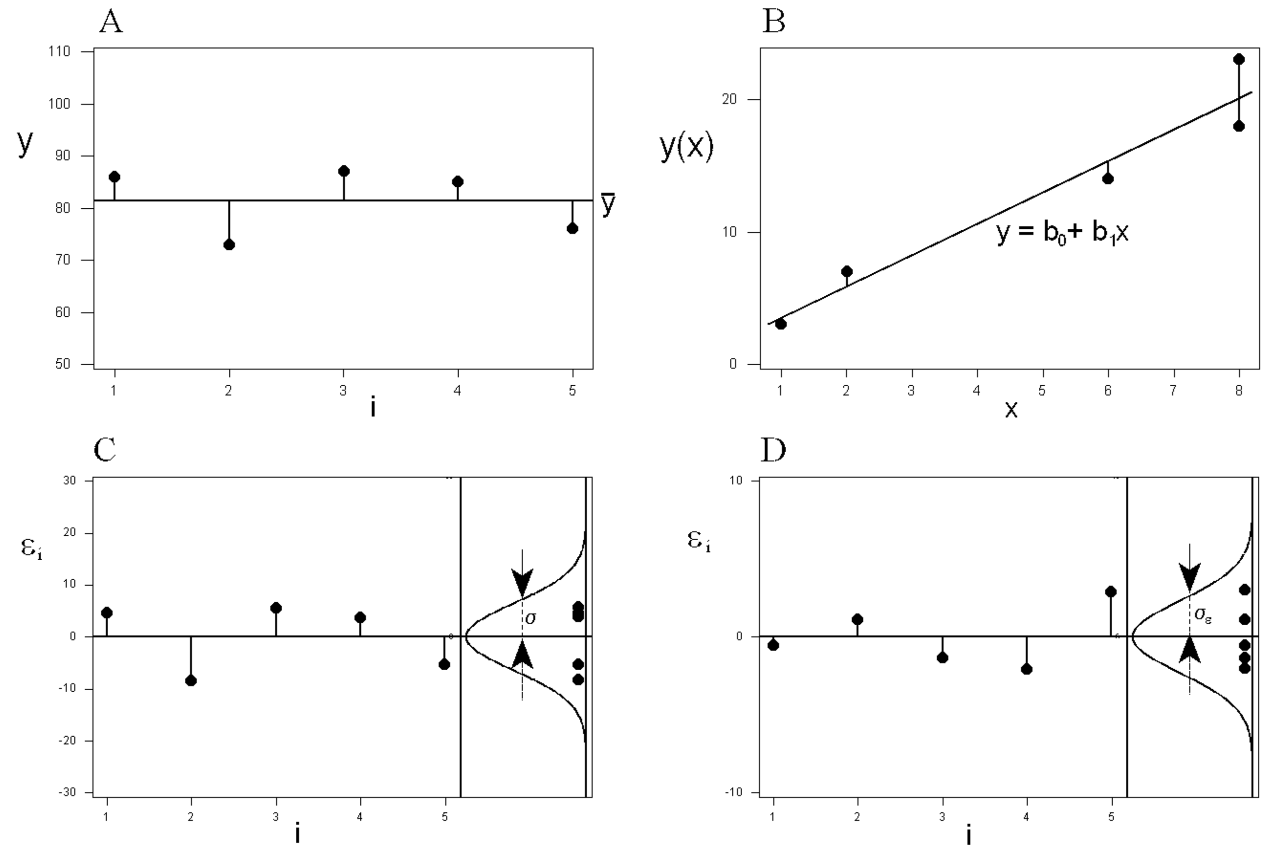
General Linear Model: Y versus P, S, WP

```
Factor  Type    Levels  Values
P       fixed        2  1, 2
WP(P)   random       6  1, 2, 3, 4, 5, 6
S       fixed        4  1, 2, 3, 4


Analysis of Variance for Y, using Adjusted SS for Tests

Source  DF   Seq SS  Adj SS  Adj MS      F      P
P        1   782.04  782.04  782.04   4.03  0.115
WP(P)    4   775.36  775.36  193.84  15.25  0.000
S        3   266.00  266.00   88.67   6.98  0.006
P*S      3    62.79   62.79   20.93   1.65  0.231
Error   12   152.52  152.52   12.71
Total   23  2038.72


S = 3.56509   R-Sq = 92.52%   R-Sq(adj) = 85.66%
```

This page is blank.

Chapter 8: Linear Regression

Compare the Models:



Method of Least Squares

The *least squares regression* line fitted to experimental data (x_i, y_i) has the form

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

where the regression coefficients b_0 and b_1 are those values that minimize the error sum of squares

$$\sum \epsilon_i^2 = \sum (y_i - \hat{y}_i)^2.$$

These values are determined from the simultaneous solution of

$$\frac{\partial}{\partial b_0} \sum \epsilon_i^2 = 0 \quad \text{and} \quad \frac{\partial}{\partial b_1} \sum \epsilon_i^2 = 0$$

which are satisfied by the line passing through point $(\bar{x}, \bar{y})$ with slope

$$b_1 = \frac{S_{xy}}{SS_x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}.$$

That is,

$$\begin{aligned} y_i &= \bar{y} + b_1 (x_i - \bar{x}) + \epsilon_i \\ &= (\bar{y} - b_1 \bar{x}) + b_1 x_i + \epsilon_i \\ &= b_0 + b_1 x_i + \epsilon_i \end{aligned}$$

Graphical Solution 1

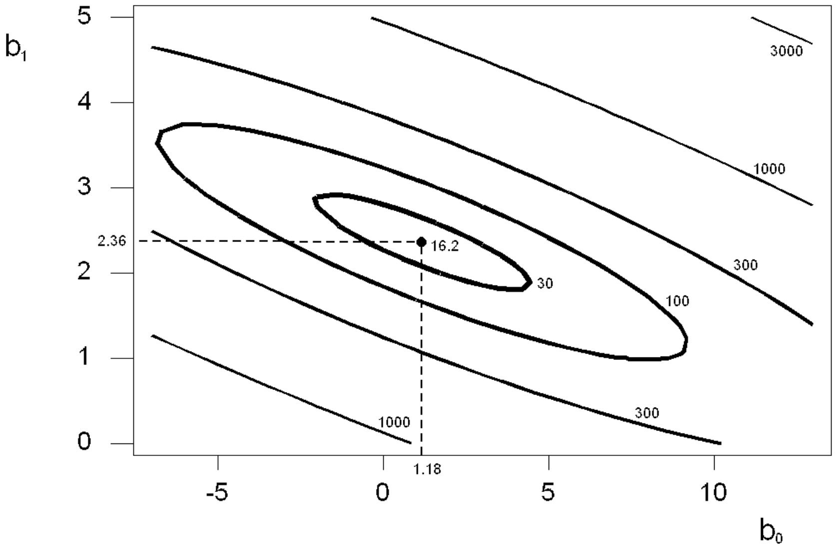
Example: A matrix of b_0 and b_1 coefficients was considered as fits to the following data:

i	1	2	3	4	5
x_i	1	2	6	8	8
y_i	3	7	14	18	23

The error sum of squares:

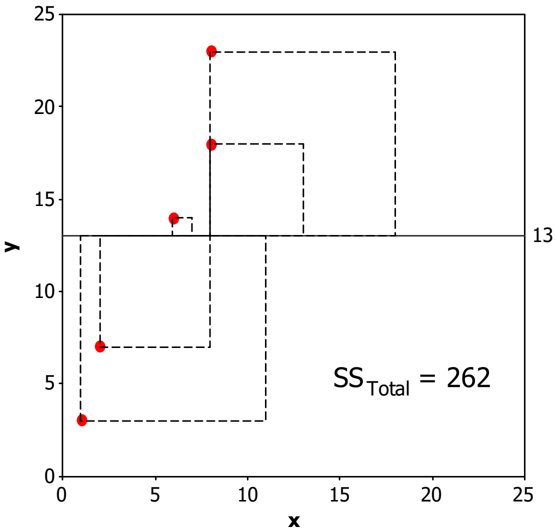
$$\sum \epsilon_i^2 = \sum (y_i - \hat{y}_i)^2$$

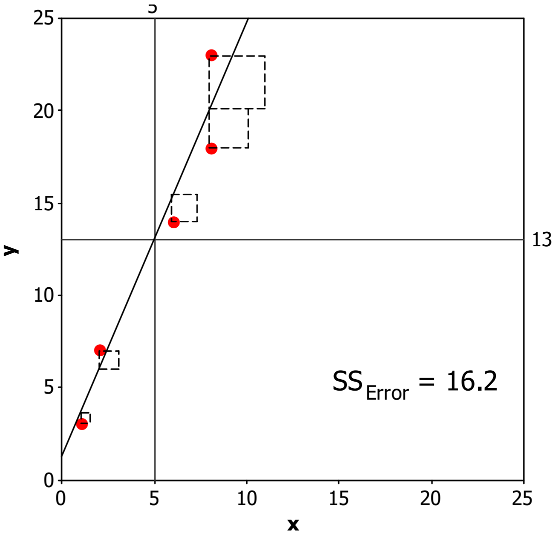
was evaluated for each (b_0, b_1) case and then the results were used to create the contour plot of $\sum \epsilon_i^2$ as a function of b_0 and b_1 shown in the following figure. Interpret the contour plot, indicate the equation of the line that provides the best fit to the data.



Graphical Solution 2

- Total variation in the response y relative to the mean $\bar{y}$ is given by SS_{Total}.
- Variation in the response relative to the least squares fitted line is given by SS_{Error}.
- Variation explained by the fitted line is given by $SS_{Regression} = SS_{Total} - SS_{Error}$





Coefficients Table for the Regression Model

Term	Coeff	SE	t	p
Constant	b_0	s_{b_0}	$t_{b_0} = b_0/s_{b_0}$	p_{b_0}
Slope	b_1	s_{b_1}	$t_{b_1} = b_1/s_{b_1}$	p_{b_1}

ANOVA Table for the Regression Model

Source	df	SS	MS	F	p
Regression	1	SS_{Regr}	$MS_{Regr} = SS_{Regr}/df_{Regr}$	$F = MS_{Regr}/MS_{Error}$	p_{Regr}
Error	$n-2$	SS_{Error}	$MS_{Error} = SS_{Error}/df_{Error}$		
Total	$n-1$	SS_{Total}			

Summary Statistics

- Standard error:

$$s_\epsilon = \sqrt{MS_{Error}}$$

- Coefficient of determination:

$$r^2 = SS_{Regr}/SS_{Total} = 1 - SS_{Error}/SS_{Total}$$

- Adjusted coefficient of determination:

$$r^2_{adj} = 1 - \frac{df_{Total}}{df_{Error}} \frac{SS_{Error}}{SS_{Total}}$$

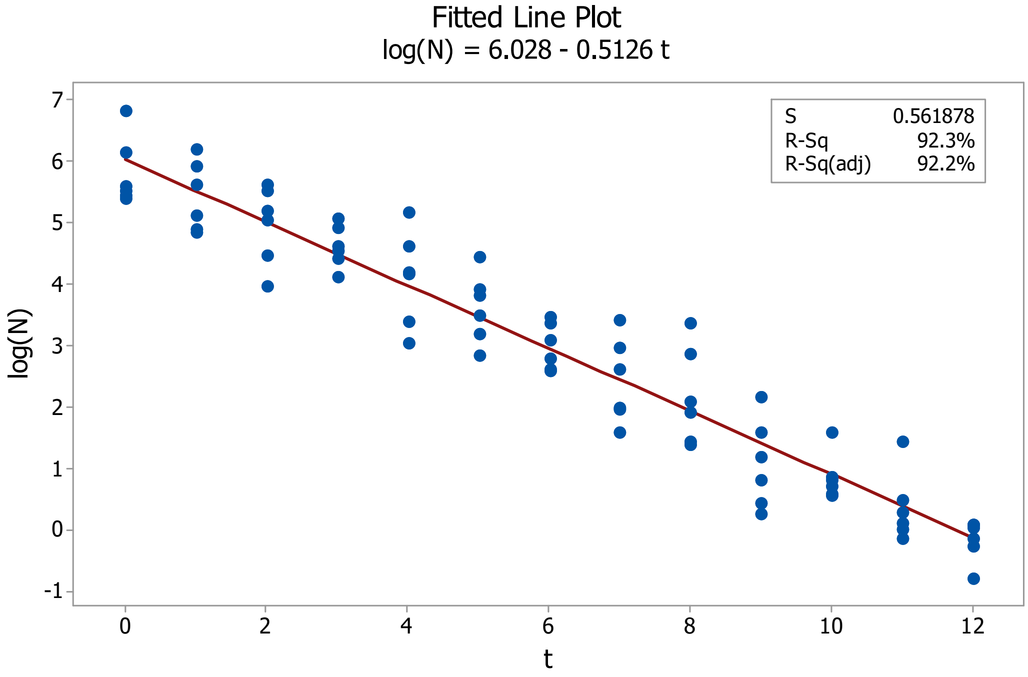
Regression Report for the Example Problem

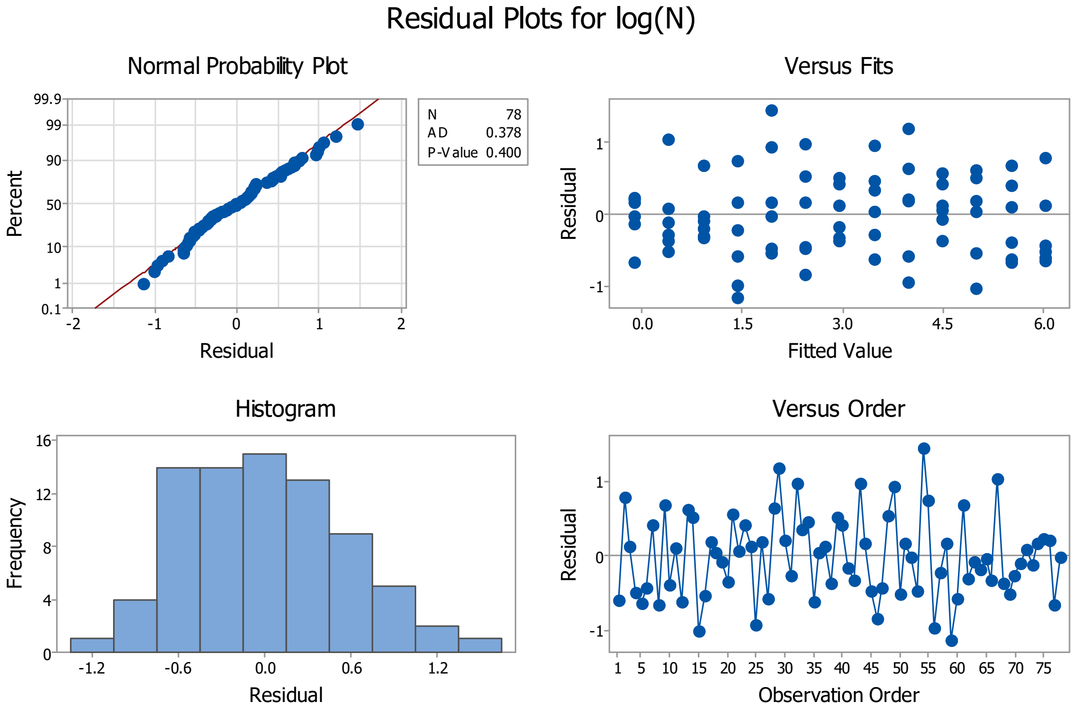
```
Regression Analysis: y versus x

Analysis of Variance

Source         DF   Adj SS   Adj MS  F-Value  P-Value
Regression      1  245.818  245.818    45.57    0.007
  x             1  245.818  245.818    45.57    0.007
Error           3   16.182    5.394
  Lack-of-Fit   2    3.682    1.841     0.15    0.879
  Pure Error    1   12.500   12.500
Total           4  262.000


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
2.32249  93.82%     91.76%      83.13%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    1.18     2.04     0.58    0.602
x          2.364    0.350     6.75    0.007  1.00


Regression Equation

y = 1.18 + 2.364 x
```

Regression Assumptions

- The x_i are known exactly, without error.
- The ϵ_i are homoscedastic with respect to the run order and the fitted values.
- The ϵ_i are normally distributed.
- The ϵ_i are independent.
- The function provides a good fit to the data.

Example:





Linear Regression with MINITAB

- Use **Stat**> **Regression**> **Fitted Line** plot to construct a scatter plot with the superimposed best fit line.
  - Turn on residuals diagnostics in the **Graph** menu.
  - Also capable of doing quadratic and cubic fits.
- Use **Stat**> **Regression**> **Regression** for a more detailed analysis.
- If the experiment has both qualitative and quantitative variables
  - (V12 to V16) Use **Stat**> **ANOVA**> **General Linear Model** and enter the quantitative variables as **Covariates**.
  - (V17) Use **Stat**> **Regression**> **Regression** or **Stat**> **ANOVA**> **General Linear Model**

Linear Regression with NCSS

Use **Analysis**> **Regression**/**Correlation**> **Linear Regression**:

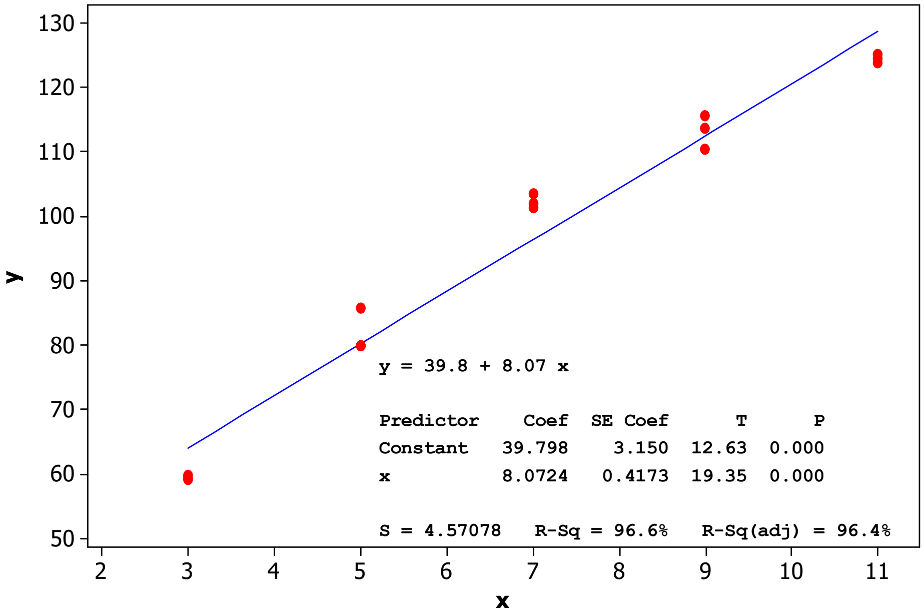
- In the **Variables** tab:
  - Specify **Y**: **Dependent Variable**.
  - Specify **X**: **Independent Variable**.
- In the **Reports** tab select:**Run Summary**, **Text Statement**, **Reg**. **Estimation**, **R2 and r**, **ANOVA**, **Assumptions**, **Y vs**. **X Plot**, **Resid**. **vs**. **X Plot**, **Histogram Plot**, **Prob**. **Plot**., and **Resid**. **vs**. **Row Plot**.
- In the **Y vs**. **X** tab turn on the **Y on X Line**, **Pred**. **Limits**, and **Confidence Limits**.

Lack of Fit or Goodness of Fit

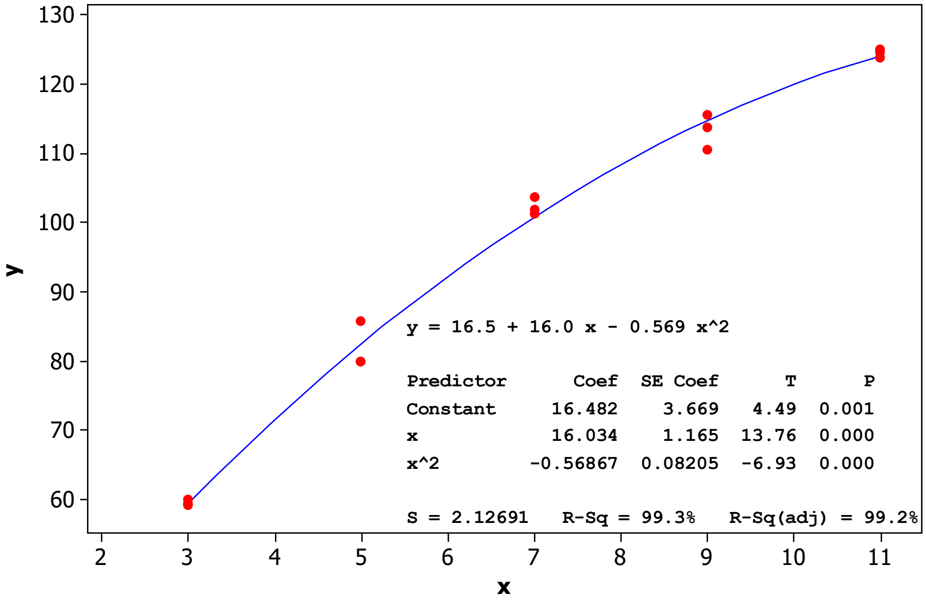
Always confirm that the linear model provides an appropriate fit to the data set using one or more of the following methods:

- Inspect the y vs. x plot with the superimposed fitted line.
- The runs test for randomness.
- Fit a quadratic model and test the quadratic regression coefficient.
- The linear lack of fit test.

Example: Although r^2 and r_{adj}^2 are very close to 1 in the following fitted line plot with linear fit, there is obviously curvature in the data. The quadratic model fitted in the next plot appears to fit the data better and the quadratic term is highly statistically significant $(p = 0.000)$. When a cubic equation was fitted to the data (not shown), the cubic regression coefficient was not statistically significant $(p = 0.585)$ so, by Occam's Razor, the cubic term may be dropped from the model.



```
y = 39.8 + 8.07 x

Predictor     Coef  SE Coef      T      P
Constant    39.798    3.150  12.63  0.000
x           8.0724   0.4173  19.35  0.000

S = 4.57078   R-Sq = 96.6%   R-Sq(adj) = 96.4%
```



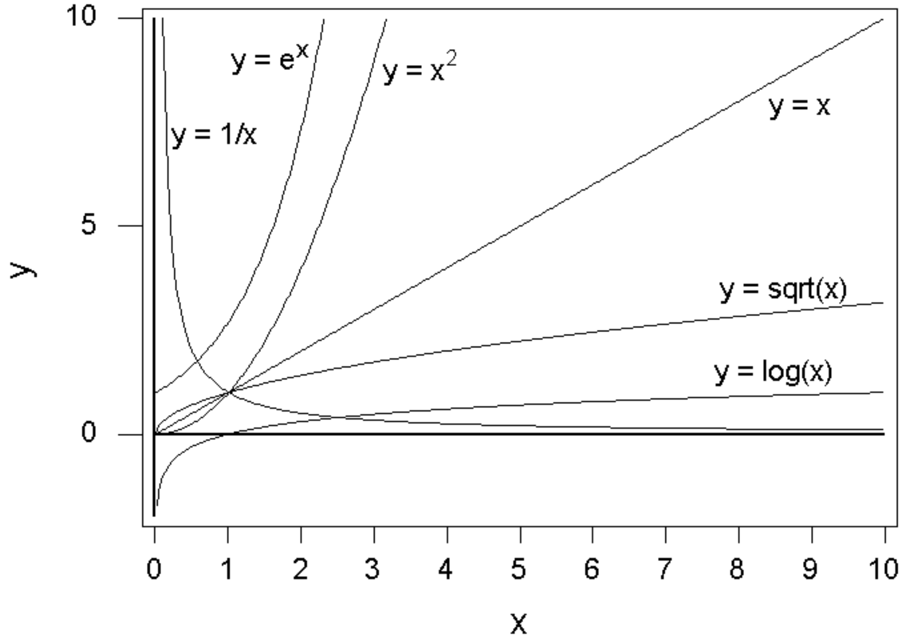
```
y = 16.5 + 16.0 x - 0.569 x^2

Predictor       Coef  SE Coef      T      P
Constant      16.482    3.669   4.49  0.001
x             16.034    1.165  13.76  0.000
x^2         -0.56867  0.08205  -6.93  0.000

S = 2.12691   R-Sq = 99.3%   R-Sq(adj) = 99.2%
```

Tansformations to Linear Form

When a linear model is not appropriate attempt a model suggested by first principles of mechanics, physics, chemistry, ...

Function	y'	x'	a'	Linear Form
$y = ae^{bx}$	$\ln y$		$\ln a$	$y' = a' + bx$
$y = ax^b$	$\log y$	$\log x$	$\log a$	$y' = a' + bx'$
$y = a + \frac{b}{x}$		$\frac{1}{x}$		$y = a + bx'$
$y = \frac{1}{a+bx}$	$\frac{1}{y}$			$y' = a + bx$
$y = ae^{\frac{b}{x}}$	$\ln y$	$\frac{1}{x}$	$\ln a$	$y' = a' + bx'$
$y = ax^2e^{bx}$	$\ln\left(\frac{y}{x^2}\right)$		$\ln a$	$y' = a' + bx$
$n = n_o e^{\frac{-\varphi}{kT}}$	$\ln n$	$\frac{1}{kT}$	$\ln n_o$	$y' = a' - \varphi x'$
$j = AT^2 e^{\frac{-\varphi}{kT}}$	$\ln\left(\frac{j}{T^2}\right)$	$\frac{1}{kT}$	$\ln A$	$y' = a' - \varphi x'$
$f(y) = a + bf(x)$	$f(y)$	$f(x)$		$y' = a + bx'$

Transformations



Finding a Variable Transformation in MINITAB and NCSS

- Use the custom MINITAB macro *%fitfinder* to create a six by six matrix of graphs of y versus x using the original, square root, square, log, power, and reciprocal transformations of both variables.
- Use NCSS's **Graphics**> **Scatter Plot Matrix**> **Functions of 2 Variables** menu to select transformations for x and y to be used in a scatter plot matrix.

Nonlinear Regression in MINITAB

Version 15:

- Method 1: Create columns for each term involving x in separate columns of the worksheet using **let** commands or the **Calc**> **Calculator** menu. Then use the **regress** command or **Stat**> **Regression**> **Regression** to perform the regression analysis by including each desired term in the model.
- Method 2: In the **Model** window of **Stat**> **ANOVA**> **General Linear Model** enter x and each desired term involving x. Enter x as a covariate so that MINITAB knows to do regression on x rather than the default choice of ANOVA.

Version 16:

- Use **Stat**> **Regression**> **Nonlinear Regression**. A catalog of common nonlinear functions is provided or you can write your own.

Nonlinear Regression in NCSS

- Create a matrix of plots with transformed x and/or y values using **Analysis**> **Curve Fitting**> **Scatter Plot Matrix**.
- Fit a user specified nonlinear function to $y(x)$ data using **Analysis**> **Curve Fitting**> **Nonlinear Regression**.

Sample Size Calculations

- Sample size can be calculated to detect a non-zero slope:

  $$H_0 : \beta_1 = 0 \text{ vs. } H_A : \beta_1 \neq 0$$

  where

  $$t = \frac{b_1}{s_{b_1}}$$

- Sample size can be calculated to determine the slope with specified values of the precision and confidence:

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

  where

  $$\delta = t_{\alpha/2} s_{b_1}$$

- Both sample size calculations involve the standard error of the regression slope:

  $$s_{b_1} = \frac{s_\epsilon}{\sqrt{SS_x}}$$

  where

  $$SS_x = \sum (x_i - \bar{x})^2$$

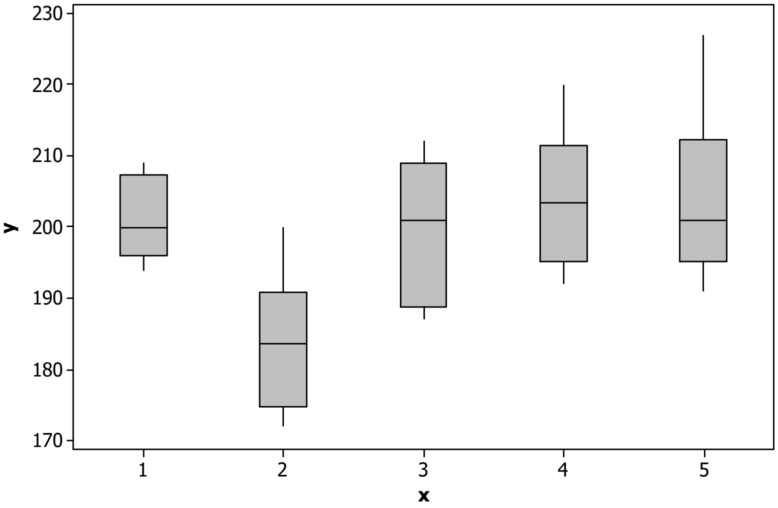
  The power of the hypothesis test or the precision of the confidence interval may be increased by increasing SS_x by:

  - Taking more observations.
  - Increasing the range of x values.
  - Concentrating observations at the end of the x interval.
- See the detailed sample size calculation instructions in Chapter 8.

ANOVA by Regression

ANOVA (with a qualitative predictor) can be performed using linear regression by creating indicator variables where each indicator variable is associated with one level of the predictor. In MINITAB use the **Calc**> **Make Indicator Variables** menu to create the columns of indicator variables and then use **Stat**> **Regression**> **Regression** with all of the indicators in the model. This is the method that MINITAB uses to analyze qualitative variables by ANOVA and quantitiative variables by regression in the **Stat**> **ANOVA**> **General Linear Model** menu; however, MINITAB hides the use of the indicator variables from the user.

Example: Analyze the data in the box plot by ANOVA and by regression.



```
General Linear Model: y versus x

Factor  Type   Levels  Values
x       fixed       5  1, 2, 3, 4, 5

Analysis of Variance for y, using Adjusted SS for Tests

Source  DF   Seq SS   Adj SS  Adj MS     F      P
x        4  2249.40  2249.40  562.35  6.30  0.001
Error   35  3124.38  3124.38   89.27
Total   39  5373.77

S = 9.44817   R-Sq = 41.86%   R-Sq(adj) = 35.21%
```

```
Regression Analysis: y versus C1=1, C1=2, C1=3, C1=4, C1=5

* C1=5 is highly correlated with other X variables
* C1=5 has been removed from the equation.


The regression equation is
y = 204 - 3.00 C1=1 - 20.1 C1=2 - 4.75 C1=3 - 0.00 C1=4

Predictor     Coef  SE Coef      T      P
Constant   204.000    3.340  61.07  0.000
C1=1        -3.000    4.724  -0.64  0.530
C1=2       -20.125    4.724  -4.26  0.000
C1=3        -4.750    4.724  -1.01  0.322
C1=4        -0.000    4.724  -0.00  1.000

S = 9.44817   R-Sq = 41.9%   R-Sq(adj) = 35.2%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       4  2249.40  562.35  6.30  0.001
Residual Error  35  3124.38   89.27
Total           39  5373.77

Source  DF   Seq SS
C1=1     1    66.31
C1=2     1  2062.76
C1=3     1   120.33
C1=4     1     0.00
```

General Linear Model

Fit $y(x,A)$ where x is a continuous predictor to be analyzed by regression (i.e. a covariate) and A is a qualitative predictor to be analyzed by ANOVA using a general linear model.

- In MINITAB use **Stat**> **ANOVA**> **General Linear Model**.
- In NCSS using **Analysis**> **ANOVA**> **GLM ANOVA**.
- Example: Fit $y(x,A)$ where x is a covariate and A has three levels 1, 2, and 3.
  - Specify the model to include the terms x, A, and $x * A$ where x is a covariate.
  - The model will have the form:

$$\begin{aligned} y_i(x,A) &= b_0 + b_{01}(A=1) + b_{02}(A=2) + b_{03}(A=3) \\ &\quad + x(b_1 + b_{11}(A=1) + b_{12}(A=2) + b_{13}(A=3)) + \epsilon_i \end{aligned}$$

  - If there are no A effects, then the model reduces to $y_i(x,A) = b_0 + b_1 x$.
  - The b_{0j} coefficients are corrections to b_0 for each level of A.
  - $b_{03} = -(b_{01} + b_{02})$
  - The b_{1j} coefficients are corrections to b_1 for each level of A.
  - $b_{13} = -(b_{11} + b_{12})$
- If y is a function of two or more covariates, avoid colinearity by mean-adjusting the covariates. For example, instead of fitting $y(x_1,x_2)$, fit $y(x_1',x_2')$ where $x_1' = x_1 - mean(x_1)$ and $x_2' = x_2 - mean(x_2)$.

Example: An experiment was performed to determine how temperature affects the growth of three different strains of tomatos. Three samples of each strain were evaluated at five different levels of temperature. Determine how the degrees of freedom are partitioned if the model must account for possible slope differences between the strains and include a generic curvature term in the model to check for lack of linear fit.

Solution: The model will have the form:

$$\begin{aligned} y &= b_0 + b_{01}(Strain=1) + b_{02}(Strain=2) + b_{03}(Strain=3) \\ &\quad + Temp[b_1 + b_{11}(Strain=1) + b_{12}(Strain=2) + b_{13}(Strain=3)] \\ &\quad + b_2 Temp^2 \end{aligned}$$

where $b_{03} = -(b_{01} + b_{02})$ and $b_{13} = -(b_{11} + b_{12})$. Note that the b_{0i} are bias corrections for the different strains and the b_{1i} are slope corrections for the different strains.

Source	*df*
Strain	2
Temp	1
Strain*Temp	2
Temp*Temp	1
Error	38
Total	44

Special Problems:

- Inverse Prediction - What is the confidence interval for the unknown x value that would be expected to deliver a specified y value?
- Errors-in-Variables - If the x values are noisy, so they are not known exactly, then the linear regression coefficients will be biased, i.e. will not correctly predict y from x. If the standard deviation of the error in x can be determined then corrected values of the regression coefficients can be calculated.
- Weighted Regression - If the residuals are not homoscedastic with respect to x_i then the observations with greater inherent noise deserve to be weighted less heavily than observations where there is less noise. If a suitable variable transformation cannot be found, then if the local variance for the observation (x_i, y_i) is σ_i^2, apply weighting factor $w_i = 1/\sigma_i^2$, i.e. (x_i, y_i, w_i).
  - In MINITAB use the weighting option in the **Options** menu of either **Stat**> **Regression**> **Regression** or **Stat**> **ANOVA**> **General Linear Model**.
  - In NCSS use the weighting option in the **Weighting Variable**: window of **Analysis**> **Regression**/**Correlation**> **Linear Regression**.
- If the response is dichotomous or binary (i.e. having just two states, e.g. pass/fail) then use binary logistic regression (BLR). In MINITAB use the **Stat**> **Regression**> **Binary Logistic Regression** menu.

Chapter 9: 2^k Experiments

Introduction

- Two levels of each of k design variables.
- Include all possible combinations of variable levels so 2^k is the number of unique runs in one replicate.
- Makes use of hidden replication.
- Can resolve main effects, two-factor interactions, and higher order interactions if desired:

$$2^k = \binom{k}{0} + \binom{k}{1} + \binom{k}{2} + \binom{k}{3} + \cdots + \binom{k}{k}$$

  We usually don't look for three-factor or higher order interactions.
- Cannot detect the presence of or quantify curvature because there are only two levels of each variable.

Coded Variables

- The two-level factorial designs are easiest to express using coded (± 1) variable levels.
- Coded levels offer significant mathematical advantages, e.g. easy to interpret variable effects.
- Coded levels add some complications, such as the need to reference the variables matrix when building the experiment from the design matrix.
- Must be able to convert back and forth between physical and coded units:
  - From physical (x) to coded (x') units:

$$x' = \frac{x - x_0}{\Delta x}$$

  - From coded to physical units:

$$x = x_0 + x'\Delta x$$

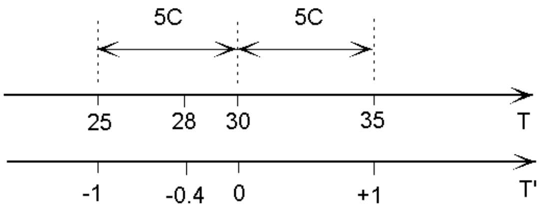
Example *An experiment is performed with two levels of temperature:* $25C$ *and* $35C$ *corresponding to coded* -1 *and* $+1$ *levels of temperature, respectively. Find the coded value that corresponds to* $28C$.

Solution*: The* 0 *level of temperature is* $x_0 = 30C$ *and the step size to the* -1 *and* $+1$ *levels is* $\Delta x = 5C$*, so the transformation equation to coded units is:*

$$x' = \frac{x - 30}{5}$$

Then the coded value of $x = 28C$ *is:*

$$x' = \frac{28 - 30}{5} = -0.4$$



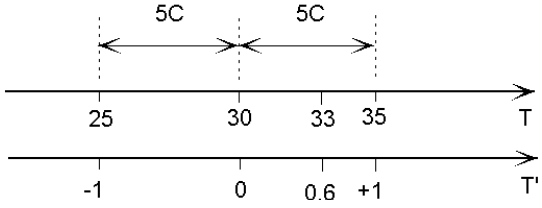
Example *Use the definitions in the preceding example to determine the temperature that has a coded value of* $x' = +0.6$.

***Solution**: The equation to transform from coded to actual values is:*

$$x = 30 + 5x'$$

so the actual temperature that corresponds to the coded value $x' = +0.6$ *is:*

$$x = 30 + 5(0.6) = 33$$



Transformation of $T' = 0.6$ Coded Units Back to

The 2^2 Experiment

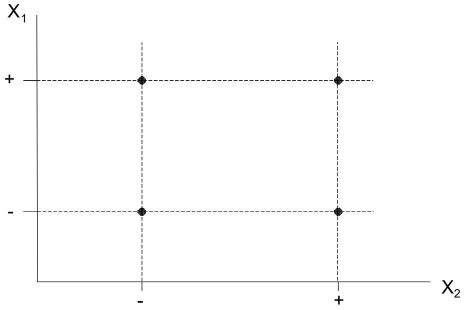
- Variables matrix:

	x_1: Temperature	x_2: Time
-1	25	3
1	35	5
Units	$°C$	min

- Design matrix:

Run	x_1	x_2
1	$-$	$-$
2	$-$	$+$
3	$+$	$-$
4	$+$	$+$

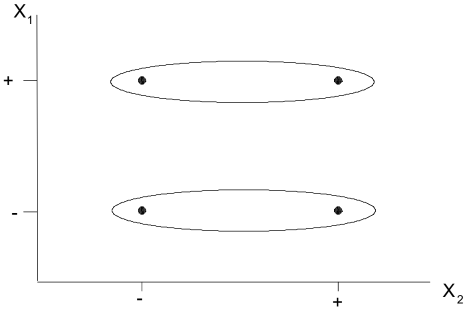
- Plot of design space in coded units:

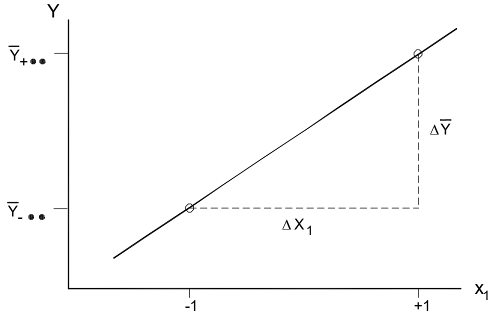


$2x2$ Factorial Design

The Effect of x_1

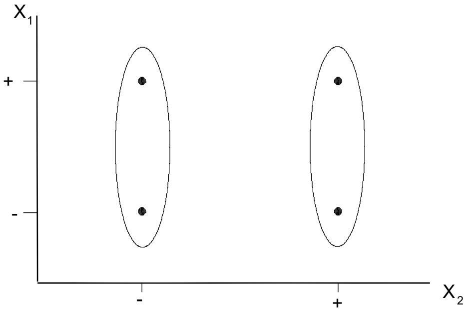
$$b_1 = \frac{\bar{y}_{+\bullet\bullet} - \bar{y}_{-\bullet\bullet}}{2}$$





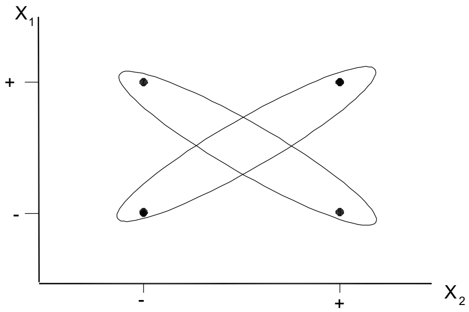
The Effect of x_2

$$b_2 = \frac{\bar{y}_{\bullet+\bullet} - \bar{y}_{\bullet-\bullet}}{2}$$



The Interaction Effect x_{12}

$$b_{12} = \frac{(\bar{y}_{++\bullet} + \bar{y}_{--\bullet}) - (\bar{y}_{+-\bullet} + \bar{y}_{-+\bullet})}{2}$$



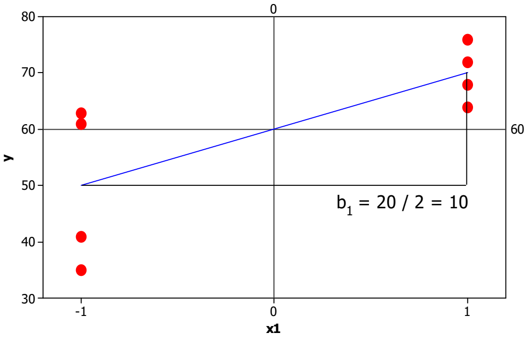
Example: Construct a model of the form:

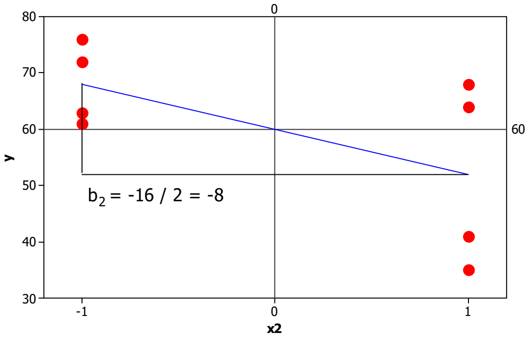
$$y(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_{12}$$

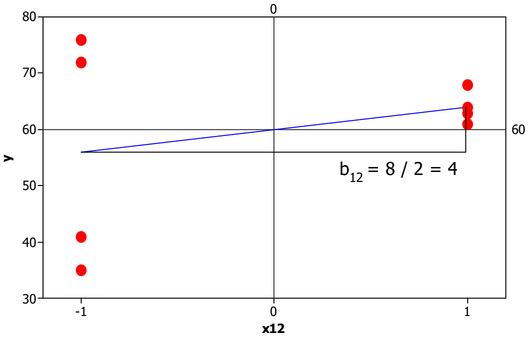
for the data set:

		x_2		
		-1	$+1$	
x_1	-1	61, 63	41, 35	$\bar{y}_{-\bullet\bullet} = 50$
	$+1$	76, 72	68, 64	$\bar{y}_{+\bullet\bullet} = 70$
		$\bar{y}_{\bullet-\bullet} = 68$	$\bar{y}_{\bullet+\bullet} = 52$	$\bar{\bar{y}}_{\bullet\bullet\bullet} = 60$

Solution:







Solution:

$$y = 60 + 10x_1 - 8x_2 + 4x_{12}$$

$Source$	b	s	t	p
$Constant$	60	1.06	57	0.00
x_1	10	1.06	9.4	0.00
x_2	−8.0	1.06	−7.5	0.00
x_{12}	4.0	1.06	3.8	0.02

$df_{total} = 7$

$df_{model} = 3$

$df_{\epsilon} = 4$

$s_{\epsilon} = 3.0$

$r^2 = 0.977$

$r^2_{adj} = 0.957$

Back to Coded Variable Levels: **Whay Are They Necessary**?

- An experiment's run matrix can be expressed in either physical/uncoded or coded units.
- When expressed in physical/uncoded units there are correlations between the terms in the model that pollute the pure values of the regression coefficients and prevent us from judging them for statistical significance.
- When expressed in coded units $(+/-1)$ the terms in the model are guaranteed to be independent so we can safely judge them for statistical significance.

Example: Consider a 2^3 design with the following variables matrix:

	A : *Temperature*	B : *Time*	C : *Pressure*
-1	25	3	40
1	35	5	60
Units	°*C*	min	psi

and run matrices in physical/uncoded and coded units:

Run	*A*	*B*	*C*
1	25	3	40
2	25	3	60
3	25	5	40
4	25	5	60
5	35	3	40
6	35	3	60
7	35	5	40
8	35	5	60

Run	*A*	*B*	*C*
1	−	−	−
2	−	−	+
3	−	+	−
4	−	+	+
5	+	−	−
6	+	−	+
7	+	+	−
8	+	+	+

The correlation matrix with the variables expressed in physical/uncoded units (created with the MINITAB macro *correlate.mac*) is:

```
             Temp   Time  Press      AB      AC      BC
     Time   0.000
     Press  0.000  0.000
     AB     0.549  0.824  0.000
     AC     0.635  0.000  0.762  0.349
     BC     0.000  0.772  0.617  0.636  0.470
     ABC    0.452  0.679  0.543  0.823  0.713  0.880
```

The correlation matrix with the variables expressed in coded units is:

```
             Temp   Time  Press      AB      AC      BC
     Time   0.000
     Press  0.000  0.000
     AB     0.000  0.000  0.000
     AC     0.000  0.000  0.000  0.000
     BC     0.000  0.000  0.000  0.000  0.000
     ABC    0.000  0.000  0.000  0.000  0.000  0.000
```

- Reputable software (e.g. MINITAB) will allow you to specify the run matrix in physical/uncoded units but will perform the analysis in coded units.
- Use MINITAB's **Stat**> **DOE**> **Display Design**> **Units for factors** menu to toggle the display of the run matrix between physical/uncoded and coded units.

Creating and Analyzing 2^k Designs in MINITAB

- Use **Stat**> **DOE**> **Factorial**> **Create Factorial Design** to create a design.
- Use **Stat**> **DOE**> **Factorial**> **Define Custom Factorial Design** to specify an existing design so that MINITAB will recognize it.
- Use **Stat**> **DOE**> **Factorial**> **Factorial Plots** to make plots of the main effects and two-factor interactions.
- Use **Stat**> **DOE**> **Factorial**> **Analyze Factorial Design** to analyze the data.
  - Enter the response in the **Responses**: window.
  - Specify the terms to be included in the model in the **Terms** window.
  - Turn on residuals diagnostic graphs and effects plots in the **Graphs** window.

Creating 2^k Designs in NCSS

Use **Analysis**> **Design of Experiments**> **Two-level Designs**:

- Specify a column for the response in **Simulated Response Variable**.
- Specify a column for blocks in **Block Variable**.
- Specify the column for the first design variable in **First Factor Variable**.
- Specify the factor levels in **Factor Values**. The values -1 and $+1$ are recommended. Specify a set of levels for as many variables as are required for the design.
- Specify the number of replicates in **Replications**:
- Specify the number of runs to be used for each block in **Block Size**.

Analyzing 2^k Experiments in NCSS

Use **Analysis**> **Design of Experiments**> **Analysis of Two-level Designs**:

- On the **Variables** tab:
  - Specify the **Response Variable**.
  - Specify the **Block Variable**.
  - Specify the **Factor Variables**.

Analyzing 2^k Experiments in NCSS

As an alternative analysis that provides more control and better residuals diagnostics use **Analysis**> **Regression**/**Correlation**> **Multiple Regression** (**2001 Edition**):

- On the **Variables** tab:
  - Specify the response in **Y**: **Dependent Variable**.
  - Specify the design variables (e.g. $A\ B\ C$) in **X's**: **Numeric Independent Variables**.
  - Specify the blocking variable in **X's**: **Categorical Independent Variables**.
- On the **Model** tab:
  - In the **Which Model Terms** window select **Custom Model**.
  - In the **Custom Model** window specify the model including block, main effects and interactions, e.g.

$$Block + A + B + C + A * B + A * C + B * C$$

- On the **Reports** tab specify: **Run Summary**, **Correlations**, **Equation**, **Coefficient**, **Write Model**, **ANOVA Summary**, **ANOVA Detail**, **Normality Tests**, **Res-X's Plots**, **Histogram**, **Probability Plot**, **Res vs Yhat Plot**, **Res vs Row Plot**.

Rules for Refining Models

- Fit the full model first, including main effects and interactions.
- Starting from the highest order interactions, begin removing the least significant ones one at a time while watching the r_{adj}^2.
- To retain an interaction in the model, all of its main effects and lower-order interactions must also be retained. For example, to retain the three-factor interaction ACE the model *must* also contain A, C, E, AC, AE, and CE.
- Don't expect to remove all of the statistically insignificant terms in the model. If the r_{adj}^2 takes a sudden plunge, put the last term back in the model.

Sample Size

The power and precision of 2^k experiments is determined by the total number of experimental runs, which is the product of the number of runs in one replicate and the number of replicates. This implies that the size of an experiment is to some degree independent of the number of variables so look for opportunities to add variables to experiments.

Sample Size to Detect an Effect

The number of experimental runs required to detect a difference δ between the ± 1 levels of a design variable with power $P = 1 - \beta$ is given by:

$$r \times 2^k \geq 4\left((t_{\alpha/2} + t_\beta) \frac{\sigma_\epsilon}{\delta} \right)^2$$

Example: An experiment is required to have 90% power $(\beta = 0.10)$ to detect an effect size of $\delta = 20$. The process is known to have $\sigma_\epsilon = 25$. How many total runs are required? How many replicates of a 2^1, 2^2, 2^3,$\ldots$ design are required?

Solution: The approximate total number of runs required is:

$$\begin{aligned} r \times 2^k &\approx 4\left((t_{\alpha/2} + t_\beta) \frac{\sigma_\epsilon}{\delta} \right)^2 \\ &\simeq 4\left((1.96 + 1.282) \frac{25}{20} \right)^2 \\ &\approx 64 \end{aligned}$$

A 2^1 design will require $64/2^1 = 32$ replicates, a 2^2 design will require $64/2^2 = 16$ replicates, a 2^3 design will require $64/2^3 = 8$ replicates, $\ldots$

Sample Size to Quantify an Effect

The number of experimental runs required to determine the regression coefficient β_i for one of the k two-level design variables with precision δ and confidence $1 - \alpha$ so that:

$$P(b_i - \delta < \beta_i < b_i + \delta) = 1 - \alpha$$

is given by:

$$r \times 2^k \geq \left(\frac{t_{\alpha/2} \sigma_\epsilon}{\delta} \right)^2$$

2^k plus Centers Design

If all k design variables are quantitative then center cells can be added to an experiment, e.g.:

x_1	x_2	x_{12}	x_{11}	x_{22}
$-$	$-$	$+$	$+$	$+$
$-$	$+$	$-$	$+$	$+$
$+$	$-$	$-$	$+$	$+$
$+$	$+$	$+$	$+$	$+$
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Center cells 1) provide extra error degrees of freedom and 2) provide a method for testing for linear lack-of-fit. The model will be of the form:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_{12}x_{12} + \cdots + b_{**}x_{**}$$

where the curvature measured by b_{**} could be due to one or more of the design variables.If the b_{**} coefficient is not statistically significant then we can remove it from the model by Occam and conclude that the simple linear model with interactions is valid. If the b_{**} coefficient is statistically and practically significant then it is necessary to perform a follow-up experiment using techniques from Chapter 11 to determine the source of the curvature. The designs from Chapter 11 can resolve the sources of curvature in a model with quadratic terms for each variable of the form:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_{12}x_{12} + \cdots + b_{11}x_1^2 + b_{22}x_2^2 + \cdots$$

Chapter 10: Fractional Factorial Experiment Designs

Motivation

2^k experiments get very large, so:

- We need a way to block large full-factorial designs.
- We don't usually need to resolve three-factor and higher order interactions.

2^k Experiments Get Very Large

If the models that we fit to 2^k experiments only include main effects and two-factor interactions, then for one replicate:

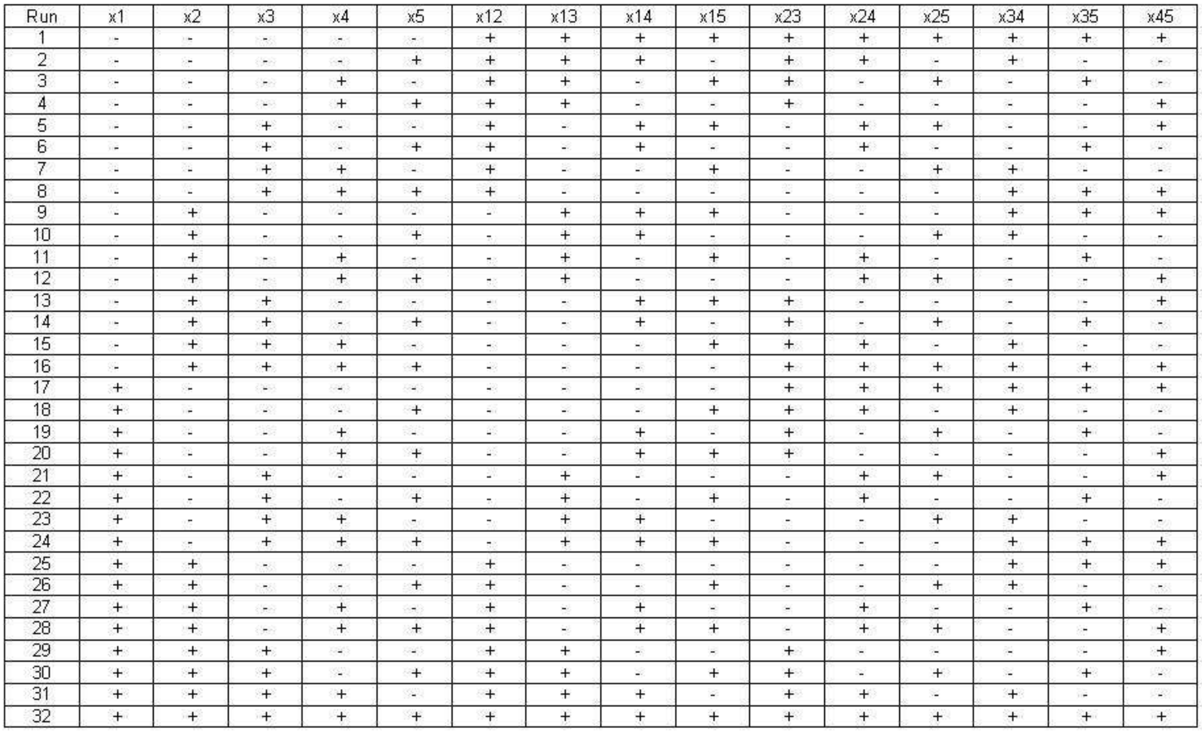
$$\begin{aligned} df_{total} &= 2^k - 1 \\ df_{model} &= \binom{k}{1} + \binom{k}{2} \\ df_{\epsilon} &= \binom{k}{3} + \cdots + \binom{k}{k} \end{aligned}$$

and df_{ϵ} increases MUCH faster than df_{model}:

k	2^k	df_{total}	df_{model}	df_{error}
2	4	3	3	0
3	8	7	6	1
4	16	15	10	5
5	32	31	15	16
6	64	63	21	42
7	128	127	28	99
8	256	255	36	219
9	512	511	45	466
10	1024	1023	55	968

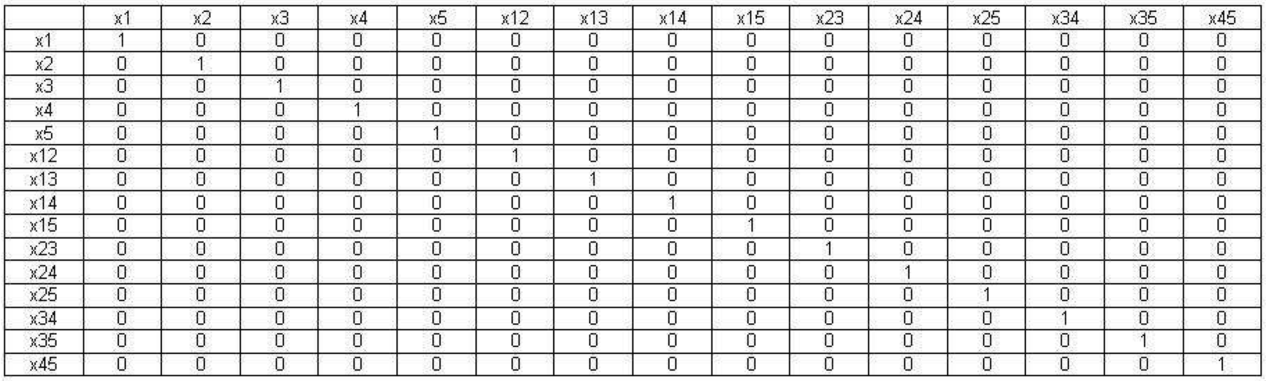
Do we really need so many error degrees of freedom?

Consider the 2^5 Design:



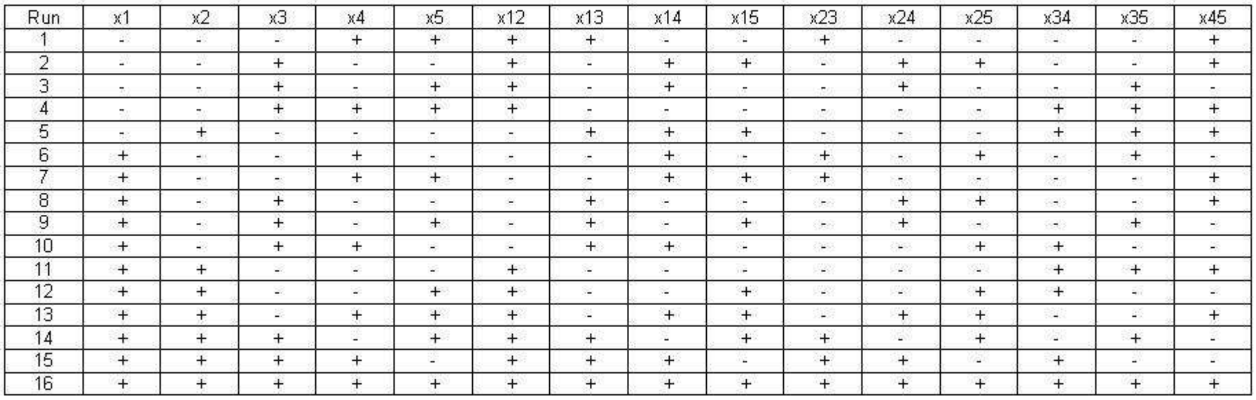
Run	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
1	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+
2	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-
3	-	-	-	+	-	+	+	-	+	+	-	+	-	+	-
4	-	-	-	+	+	+	+	-	-	+	-	-	-	-	+
5	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
6	-	-	+	-	+	+	-	+	-	-	+	-	-	+	-
7	-	-	+	+	-	+	-	-	+	-	-	+	+	-	-
8	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
9	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
10	-	+	-	-	+	-	+	+	-	-	-	+	+	-	-
11	-	+	-	+	-	-	+	-	+	-	+	-	-	+	-
12	-	+	-	+	+	-	+	-	-	-	+	+	-	-	+
13	-	+	+	-	-	-	-	+	+	+	-	-	-	-	+
14	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
15	-	+	+	+	-	-	-	-	+	+	+	-	+	-	-
16	-	+	+	+	+	-	-	-	-	+	+	+	+	+	+
17	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+
18	+	-	-	-	+	-	-	-	+	+	+	-	+	-	-
19	+	-	-	+	-	-	-	+	-	+	-	+	-	+	-
20	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
21	+	-	+	-	-	-	+	-	-	-	+	+	-	-	+
22	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
23	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
24	+	-	+	+	+	-	+	+	+	-	-	-	+	+	+
25	+	+	-	-	-	+	-	-	-	-	-	-	+	+	+
26	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
27	+	+	-	+	-	+	-	+	-	-	+	-	-	+	-
28	+	+	-	+	+	+	-	+	+	-	+	+	-	-	+
29	+	+	+	-	-	+	+	-	-	+	-	-	-	-	+
30	+	+	+	-	+	+	+	-	+	+	-	+	-	+	-
31	+	+	+	+	-	+	+	+	-	+	+	-	+	-	-
32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

The correlation matrix for the 2^5 design:



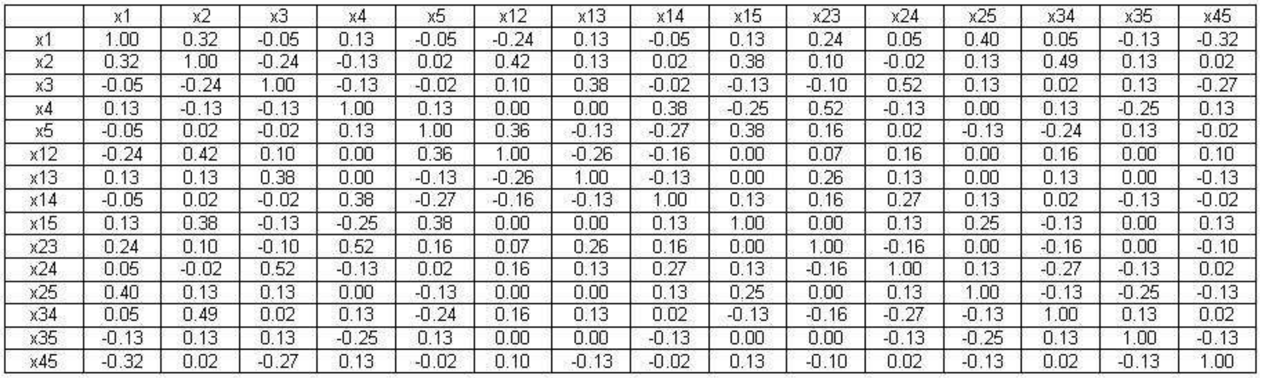
	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
x1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
x3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
x4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
x5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
x12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x15	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x23	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x24	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
x25	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
x34	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
x35	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
x45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Suppose That We Use a Random 16 Run Subset:



Run	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
1	-	-	-	+	+	+	+	-	-	+	-	-	-	-	+
2	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
3	-	-	+	-	+	+	-	+	-	-	+	-	-	+	-
4	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
5	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
6	+	-	-	+	-	-	-	+	-	+	-	+	-	+	-
7	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
8	+	-	+	-	-	-	+	-	-	-	+	+	-	-	+
9	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
10	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
11	+	+	-	-	-	+	-	-	-	-	-	-	+	+	+
12	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
13	+	+	-	+	+	+	-	+	+	-	+	+	-	-	+
14	+	+	+	-	+	+	+	-	+	+	-	+	-	+	-
15	+	+	+	+	-	+	+	+	-	+	+	-	+	-	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Correlation matrix for the experiment of 16 randomly chosen runs:

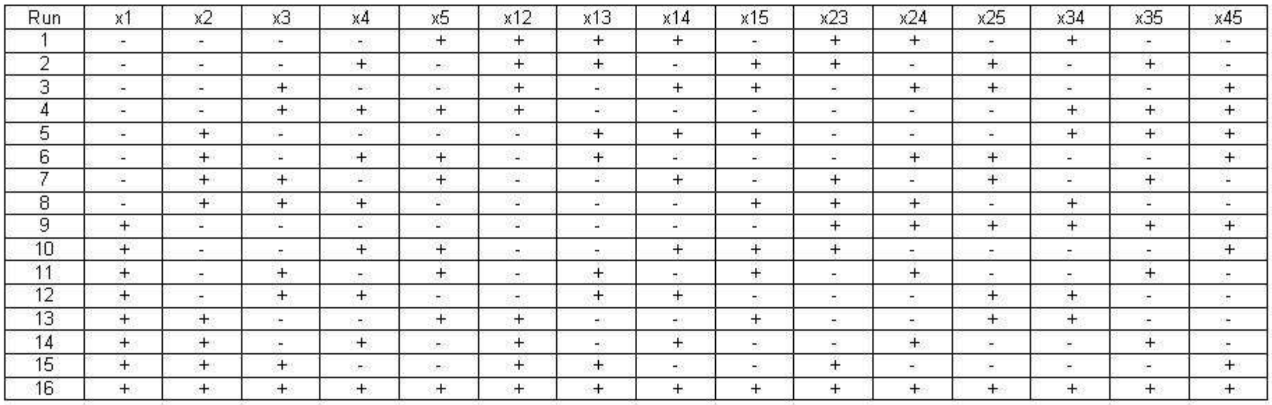


	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
x1	1.00	0.32	-0.05	0.13	-0.05	-0.24	0.13	-0.05	0.13	0.24	0.05	0.40	0.05	-0.13	-0.32
x2	0.32	1.00	-0.24	-0.13	0.02	0.42	0.13	0.02	0.38	0.10	-0.02	0.13	0.49	0.13	0.02
x3	-0.05	-0.24	1.00	-0.13	-0.02	0.10	0.38	-0.02	-0.13	-0.10	0.52	0.13	0.02	0.13	-0.27
x4	0.13	-0.13	-0.13	1.00	0.13	0.00	0.00	0.38	-0.25	0.52	-0.13	0.00	0.13	-0.25	0.13
x5	-0.05	0.02	-0.02	0.13	1.00	0.36	-0.13	-0.27	0.38	0.16	0.02	-0.13	-0.24	0.13	-0.02
x12	-0.24	0.42	0.10	0.00	0.36	1.00	-0.26	-0.16	0.00	0.07	0.16	0.00	0.16	0.00	0.10
x13	0.13	0.13	0.38	0.00	-0.13	-0.26	1.00	-0.13	0.00	0.26	0.13	0.00	0.13	0.00	-0.13
x14	-0.05	0.02	-0.02	0.38	-0.27	-0.16	-0.13	1.00	0.13	0.16	0.27	0.13	0.02	-0.13	-0.02
x15	0.13	0.38	-0.13	-0.25	0.38	0.00	0.00	0.13	1.00	0.00	0.13	0.25	-0.13	0.00	0.13
x23	0.24	0.10	-0.10	0.52	0.16	0.07	0.26	0.16	0.00	1.00	-0.16	0.00	-0.16	0.00	-0.10
x24	0.05	-0.02	0.52	-0.13	0.02	0.16	0.13	0.27	0.13	-0.16	1.00	0.13	-0.27	-0.13	0.02
x25	0.40	0.13	0.13	0.00	-0.13	0.00	0.00	0.13	0.25	0.00	0.13	1.00	-0.13	-0.25	-0.13
x34	0.05	0.49	0.02	0.13	-0.24	0.16	0.13	0.02	-0.13	-0.16	-0.27	-0.13	1.00	0.13	0.02
x35	-0.13	0.13	0.13	-0.25	0.13	0.00	0.00	-0.13	0.00	0.00	-0.13	-0.25	0.13	1.00	-0.13
x45	-0.32	0.02	-0.27	0.13	-0.02	0.10	-0.13	-0.02	0.13	-0.10	0.02	-0.13	0.02	-0.13	1.00

There are $\binom{32}{16} = 601,000,000$ different 16-run subsets. Most of them will have undesireable correlation matrices, but some will not.

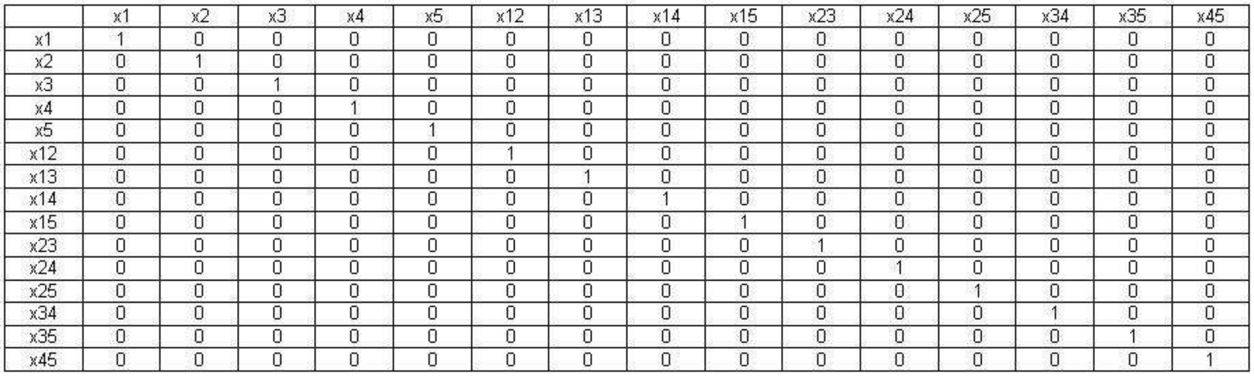
If We Can’t Beat the Correlations, Can We at Least Find a Way to Tolerate Them?

Consider only those runs where $x_5 = x_1 x_2 x_3 x_4 = x_{1234}$:



Run	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
1	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-
2	-	-	-	+	-	+	+	-	+	+	-	+	-	+	-
3	-	-	+	-	-	+	-	+	+	-	+	+	-	-	+
4	-	-	+	+	+	+	-	-	-	-	-	-	+	+	+
5	-	+	-	-	-	-	+	+	+	-	-	-	+	+	+
6	-	+	-	+	+	-	+	-	-	-	+	+	-	-	+
7	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
8	-	+	+	+	-	-	-	-	+	+	+	-	+	-	-
9	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+
10	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
11	+	-	+	-	+	-	+	-	+	-	+	-	-	+	-
12	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-
13	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-
14	+	+	-	+	-	+	-	+	-	-	+	-	-	+	-
15	+	+	+	-	-	+	+	-	-	+	-	-	-	-	+
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Correlation matrix for the 16 run experiment with $x_5 = x_{1234}$:



	x1	x2	x3	x4	x5	x12	x13	x14	x15	x23	x24	x25	x34	x35	x45
x1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
x3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
x4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
x5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
x12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
x13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
x14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
x15	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
x23	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
x24	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
x25	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
x34	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
x35	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
x45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

This experiment contains one half of the original 32 run 2^5 full-factorial design so it is designated a 2^{5-1} half-fractional factorial design.

How Was This Design Determined?

$$x_5 = x_1 x_2 x_3 x_4 = x_{1234}$$

or

$$5 = 1234$$

and this implies:

$1 = 2345$	$12 = 345$	$23 = 145$	$34 = 125$	$45 = 123$
$2 = 1345$	$13 = 245$	$24 = 135$	$35 = 124$	
$3 = 1245$	$14 = 235$	$25 = 134$		
$4 = 1235$	$15 = 234$			
$5 = 1234$				

For example:

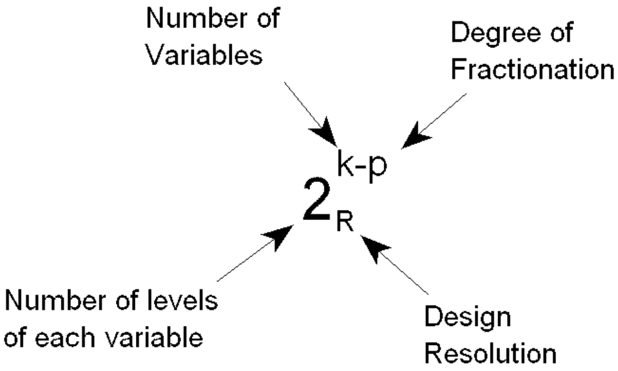
$$15 = 11234 = (11)234 = 234$$

Design Resolution

- In a fractional factorial design, every confounding relation contains the same number of variables. (This is not quite true, but for the moment...)
- The number of variables in a confounding relation is called the design resolution.
- The design designation, e.g. 2^{5-1}, is modified by adding a Roman numeral subscript, e.g. V, IV, III, to indicate the design resolution.
- Example: The 2^{5-1} design confounds main effects with four factor interactions (e.g. $5 = 1234$) and two-factor interactions with three-factor interactions (e.g. $12 = 345$) so the design is Resolution V:

$$2_V^{5-1}$$

Design Designation



Analysis of the 2_V^{5-1} Saturated Design

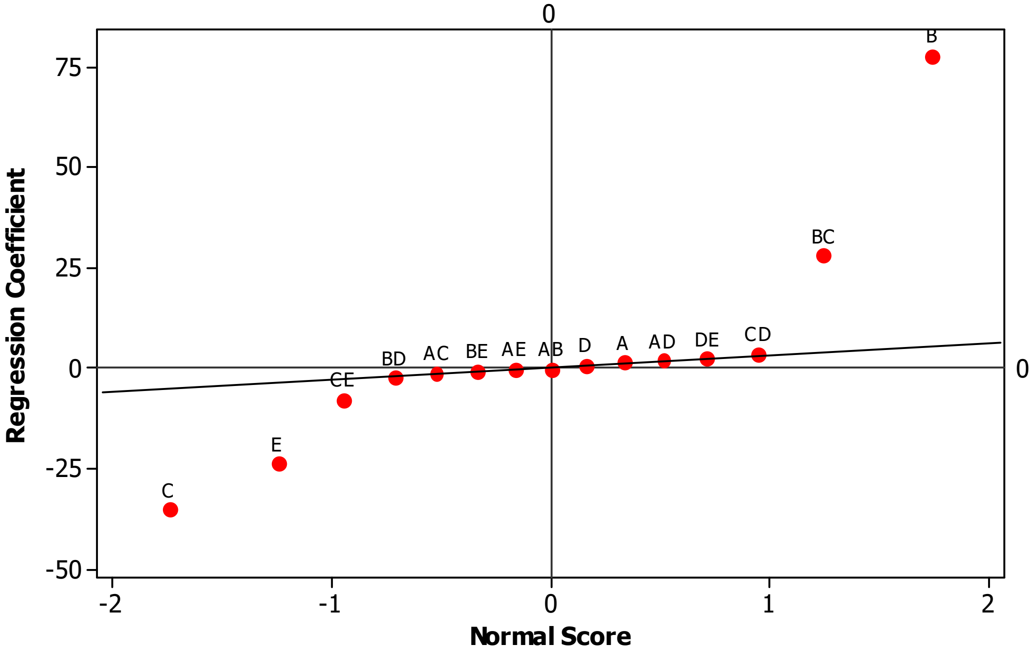
- In the resolution V design, we must assume that all three-factor and higher order interactions are insignificant so the model contains only main effects and two-factor interactions. This model consumes $df_{model} = 5 + 10 = 15$ degrees of freedom.

$$
\begin{aligned}
y = b_0 &+ b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 \\
&+ b_{12}x_{12} + b_{13}x_{13} + b_{14}x_{14} + b_{15}x_{15} \\
&+ b_{23}x_{23} + b_{24}x_{24} + b_{25}x_{25} \\
&+ b_{34}x_{34} + b_{35}x_{35} \\
&+ b_{45}x_{45}
\end{aligned}
$$

- If an experiment uses only one replicate of the 2_V^{5-1} design, the model will consume all available degrees of freedom:

$$df_\epsilon = df_{total} - df_{model} = 15 - 15 = 0$$

  Such designs are called *saturated* designs.
- To analyze a saturated design either:
  - Use an independent estimate of σ_ϵ to construct the required F tests.
  - Fit the model with main effects and two-factor interactions and construct the normal probability plot of the regression coefficients. Many of the regression coefficients can be expected to be negligible $(b_i \simeq 0)$ and will fall on an approximately straight line near the center of the normal plot. Any outlying coefficients are possibly significant. Use a reverse stepwise algorithm to refine the model by dropping the weakest model terms first.



Fractional Factorial Designs and Generators

k	Design Resolution	Design	Runs	Generators
3	III	2_{III}^{3-1}	4	$3 = \pm 12$
4	IV	2_{IV}^{4-1}	8	$4 = \pm 123$
5	III	2_{III}^{5-2}	8	$4 = \pm 12,\ 5 = \pm 13$
	V	2_{V}^{5-1}	16	$5 = \pm 1234$
6	III	2_{III}^{6-3}	8	$4 = \pm 12,\ 5 = \pm 13,\ 6 = \pm 23$
	IV	2_{IV}^{6-2}	16	$5 = \pm 123,\ 6 = \pm 234$
	VI	2_{VI}^{6-1}	32	$6 = \pm 12345$
7	III	2_{III}^{7-4}	8	$4 = \pm 12,\ 5 = \pm 13,\ 6 = \pm 23,\ 7 = \pm 123$
	IV	2_{IV}^{7-3}	16	$5 = \pm 123,\ 6 = \pm 234,\ 7 = \pm 134$
	IV	2_{IV}^{7-2}	32	$6 = \pm 1234,\ 7 = \pm 1245$
	VII	2_{VII}^{7-1}	64	$7 = \pm 123456$
8	IV	2_{IV}^{8-4}	16	$5 = \pm 234,\ 6 = \pm 134,\ 7 = \pm 123,\ 8 = \pm 124$
	IV	2_{IV}^{8-3}	32	$6 = \pm 123,\ 7 = \pm 124,\ 8 = \pm 2345$
	V	2_{V}^{8-2}	64	$7 = \pm 1234,\ 8 = \pm 1256$
	$VIII$	2_{VIII}^{8-1}	128	$8 = \pm 1234567$

The 2_{IV}^{4-1} Design

- The design generator is:

$$4 = \pm 123$$

- The confounding relations are:

$1 = 234$	$12 = 34$
$2 = 134$	$13 = 24$
$3 = 124$	$14 = 23$
$4 = 123$	

- All confounding relations include 4 variables so the design is Resolution IV:

$$2_{IV}^{4-1}$$

- Determine the matrix of runs by starting from the 2^3 design in 8 runs and generate x_4 with the design generator.

The 2_{IV}^{4-1} Design

Run matrix for the 2_{IV}^{4-1} Design

Run	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_{12}$	$\mathbf{x}_{13}$	$\mathbf{x}_{14}$	$\mathbf{x}_{23}$	$\mathbf{x}_{24}$	$\mathbf{x}_{34}$
1	-	-	-	-	+	+	+	+	+	+
2	-	-	+	+	+	-	-	-	-	+
3	-	+	-	+	-	+	-	-	+	-
4	-	+	+	-	-	-	+	+	-	-
5	+	-	-	+	-	-	+	+	-	-
6	+	-	+	-	-	+	-	-	+	-
7	+	+	-	-	+	-	-	-	-	+
8	+	+	+	+	+	+	+	+	+	+

Correlation matrix for the 2_{IV}^{4-1} Design

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_{12}$	$\mathbf{x}_{13}$	$\mathbf{x}_{14}$	$\mathbf{x}_{23}$	$\mathbf{x}_{24}$	$\mathbf{x}_{34}$
$\mathbf{x}_1$	1	0	0	0	0	0	0	0	0	0
$\mathbf{x}_2$	0	1	0	0	0	0	0	0	0	0
$\mathbf{x}_3$	0	0	1	0	0	0	0	0	0	0
$\mathbf{x}_4$	0	0	0	1	0	0	0	0	0	0
$\mathbf{x}_{12}$	0	0	0	0	1	0	0	0	0	1
$\mathbf{x}_{13}$	0	0	0	0	0	1	0	0	1	0
$\mathbf{x}_{14}$	0	0	0	0	0	0	1	1	0	0
$\mathbf{x}_{23}$	0	0	0	0	0	0	1	1	0	0
$\mathbf{x}_{24}$	0	0	0	0	0	1	0	0	1	0
$\mathbf{x}_{34}$	0	0	0	0	1	0	0	0	0	1

Analysis of the 2_{IV}^{4-1} Design

- The model for the 2^4 full factorial design can include all possible terms:

$$\begin{aligned} y = {} & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \\ & + b_{12} x_{12} + b_{13} x_{13} + b_{14} x_{14} + b_{23} x_{23} + b_{24} x_{24} + b_{34} x_{34} \\ & + b_{123} x_{123} + b_{124} x_{124} + b_{134} x_{134} + b_{234} x_{234} \\ & + b_{1234} x_{1234} \end{aligned}$$

- We cannot include all of those terms the model for the 2_{IV}^{4-1} design:

$$\begin{aligned} y = {} & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \\ & + b_{12} x_{12} + b_{13} x_{13} + b_{14} x_{14} \end{aligned}$$

  because $x_{12} = x_{34}$, $x_{13} = x_{24}$, and $x_{14} = x_{23}$.
- Use Occam and follow-up experiments to interpret the significant interaction terms.

Example: A 2_{IV}^{4-1} experiment yields the following model. The significant coefficients are indicated with an "*". Simplify the model.

$$\begin{aligned} y = {} & b_0^* + b_1 x_1 + b_2^* x_2 + b_3^* x_3 + b_4 x_4 \\ & + b_{12} x_{12} + b_{13} x_{13} + b_{14}^* x_{14} \end{aligned}$$

Solution: The x_{14} term is probably not the true source of the effect because x_1 and x_4 are not significant. But x_{14} is confounded with x_{23}. It is much more likely that x_{23} is the real source of the effect since x_2 and x_3 are both significant. The model reduces to:

$$y = b_0^* + b_2^* x_2 + b_3^* x_3 + b_{23}^* x_{23}$$

The Consequences of Confounding

- If $12 = 34$ then $b_{12}^{(full)} + b_{34}^{(full)} \rightarrow b_{12}^{(fractional)}$
- If $12 = -34$ then $b_{12}^{(full)} - b_{34}^{(full)} \rightarrow b_{12}^{(fractional)}$
- Two insignificant terms in the full design can add to become marginally significant in the fractional design:

$$b_{12} + b_{34} \rightarrow b_{12}^{*}$$

- Two significant terms in the full design can cancel out to become insignificant in the fractional design:

$$b_{12}^{*} + b_{34}^{*} \rightarrow b_{12}$$

More Highly Fractionated Designs (2^{k-p})

- 2^{k-1} is a half fractional factorial design.
- 2^{k-2} is a quarter fractional factorial design.
- 2^{k-3} is an eighth fractional factorial design.
- 2^{k-4} is a sixteenth fractional factorial design.
- If the design is 2^{k-p} then there will be p generators.

The 2_{III}^{7-4} Design

- Start from a 2^3 design with 8 runs.
- The generators for variables x_4, x_5, x_6, and x_7 are:

$$\begin{aligned} x_4 &= x_{12} \\ x_5 &= x_{13} \\ x_6 &= x_{23} \\ x_7 &= x_{123} \end{aligned}$$

- The shortest generator/confounding relation has three variables so this is a Resolution III design.
- Since all main effects are confounded with two-factor interactions we must assume that the interactions are not significant so:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7$$

Analyzing the 2_{III}^{3-1} Design

- The confounding relations are:

$$\begin{aligned} x_1 &= x_{23} \\ x_2 &= x_{13} \\ x_3 &= x_{12} \end{aligned}$$

- We can only include main effects in the model:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

- But is the model with main effects correct, or is one of the following models the right one?

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_{12}$$

$$y = b_0 + b_1x_1 + b_3x_3 + b_{13}x_{13}$$

$$y = b_0 + b_2x_2 + b_3x_3 + b_{23}x_{23}$$

Folding

- Two folded Resolution *III* designs always form a Resolution *IV* design.
- Fold an experiment by inverting all of the + and − variable levels.
- Run the original Resolution *III* design and its fold-over in separate blocks.
- Analyze them together for main effects and select two-factor interactions.
- Folding can be also be used with higher resolution designs. For example, the fold-over of a half-fractional factorial design is just the complementary half-fraction to the original design.

Use of Fractional Factorial Designs

- Avoid the use of resolution III designs except to define blocks in designs of higher resolution.
- Resolution IV designs occasionally provide enough information to answer general questions.
- Use resolution IV designs to define blocks in designs of higher resolution.
- Resolution V designs are considered safe.

Creating and Analyzing 2^{k-p} Designs in MINITAB

Use the same tools to design and analyze fractional factorial designs in MINITAB as are used for full factorial designs.

- Use **Stat**> **DOE**> **Factorial**> **Create Factorial Design** to create a design.
- Use **Stat**> **DOE**> **Factorial**> **Define Custom Factorial Design** to specify an existing design so that MINITAB will recognize it.
- Use **Stat**> **DOE**> **Factorial**> **Factorial Plots** to make plots of the main effects and two-factor interactions.
- Use **Stat**> **DOE**> **Factorial**> **Analyze Factorial Design** to analyze the data.
  - Enter the response in the **Responses**: window.
  - Specify the terms to be included in the model in the **Terms** window. When refining a model, it may be necessary to remove an interaction from a model and replace it with another interaction that the first is confounded with. For example, if $AB = CD$ and the original model shows that A, B, and CD are statistically significant, then replace CD with AB.
  - Turn on residuals diagnostic graphs and effects plots in the **Graphs** window.
- Use **Stat**> **DOE**> **Modify Design**> **Fold Design** to fold the original design.

Creating and Analyzing 2^{k-p} Designs in NCSS

Create a fractional factorial experiment using **Analysis**> **Design of Experiments**> **Fractional Factorial Designs**:

- Specify a column for the response in **Simulated Response Variable** (e.g. c1 or Y).
- Specify a column for blocks in **Block Variable** (e.g. c2 or Blocks).
- Specify the column for the first design variable in **First Factor Variable** (e.g. c3 or A)
- Specify the factor levels in **Factor Values**. The values -1 and $+1$ are recommended. Specify a set of levels for as many variables as are required for the design.
- Specify the number of experimental runs in **Runs**.
- Specify the number of runs to be used for each block in **Block Size**.

Analyze the experiment using **Analysis**> **Design of Experiments**> **Analysis of Two-level Designs** or **Analysis**> **Regression/Correlation**> **Multiple Regression** (**2001 Edition**). See the notes from Chapter 9 for details for configuring these analyses.

Plackett-Burman Designs

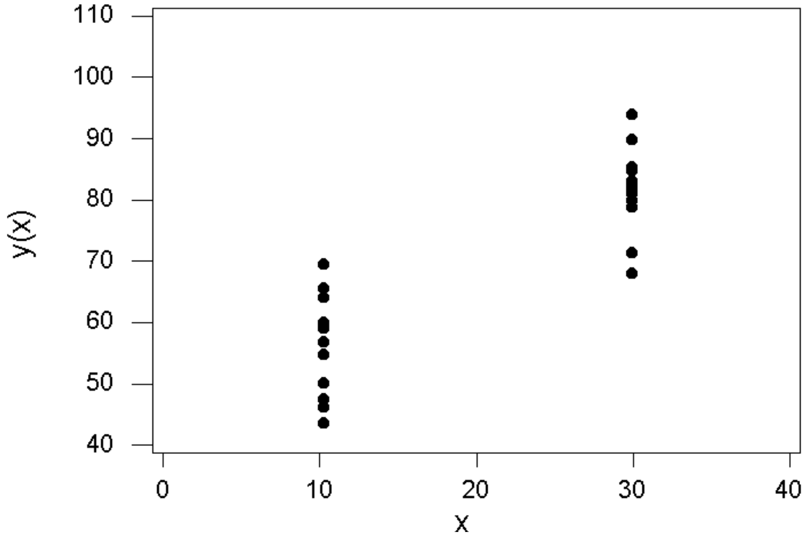
- Plackett-Burman (P-B) designs are a special form of highly fractionated two-level designs.
- All P-B designs are resolution III, i.e. main effects are confounded with two-factor interactions; however, the correlations between the main effects and two-factor interactions are less than one with the exception of the 8 run design.
- If A is confounded with BC, BD, etc., then $b_A^{fractional} = b_A^{full} + r_{A,BC} b_{BC}^{full} + r_{A,BD} b_{BD}^{full} + \cdots$
- P-B designs are primarily used for screening experiments and robust design validation studies.
- P-B designs have N runs where N is a multiple of 4, so there are P-B designs for 4, 8, 12, 16, 20, ... runs.
- The P-B designs are redundant with the 2^{k-p} designs when 2^{k-p} is an integer multiple of 4, i.e. those designs with 4, 8, 16, 32, ... runs
- P-B designs can resolve up to $N - 1$ main effects.
- If an experiment has less than $N - 1$ variables, then just leave the extra variables out of the model, i.e. pool them with the error estimate.
- With respect to every pair of variables, e.g. A and B, the experiment collapses to a 2^2 design with replicates.
- Every variable is confounded with two-factor interactions involving all other variables except itself, e.g. A will be confounded with two-factor interactions involving B, C, ... but none involving A.
- The P-B design generator is the first row of the design matrix. The other rows are generated by shifting the signs by one position for each successive row and finally adding an Nth row of all minus signs to preserve the design's balance.
- Example: 12 run P-B design with 11 design variables in standard order:

Run	A	B	C	D	E	F	G	H	J	K	L
1	+	−	+	−	−	−	+	+	+	−	+
2	+	+	−	+	−	−	−	+	+	+	−
3	−	+	+	−	+	−	−	−	+	+	+
4	+	−	+	+	−	+	−	−	−	+	+
5	+	+	−	+	+	−	+	−	−	−	+
6	+	+	+	−	+	+	−	+	−	−	−
7	−	+	+	+	−	+	+	−	+	−	−
8	−	−	+	+	+	−	+	+	−	+	−
9	−	−	−	+	+	+	−	+	+	−	+
10	+	−	−	−	+	+	+	−	+	+	−
11	−	+	−	−	−	+	+	+	−	+	+
12	−	−	−	−	−	−	−	−	−	−	−

- Create the fold-over design of a P-B design by inverting all of the +/- signs in the original design matrix. Use the custom MINITAB macro *fold.mac* to append the fold-over design to the original P-B design.
- As with other resolution III designs, the P-B design combined with its fold-over is resolution IV. Such designs provide VERY USEFUL screening experiments for processes with many variables. These designs have considerable confounding between two-factor interactions but provide excellent resolution for main effects - meeting the goal of the design for screening experiments.
- Example: The 12 run P-B design combined with its 12 run fold-over, giving a total of 24 runs, is resolution IV so can resolve up to 11 main effects (confounded with three factor interactions) and 11 two-factor interactions (confounded with other two-factor interactions).

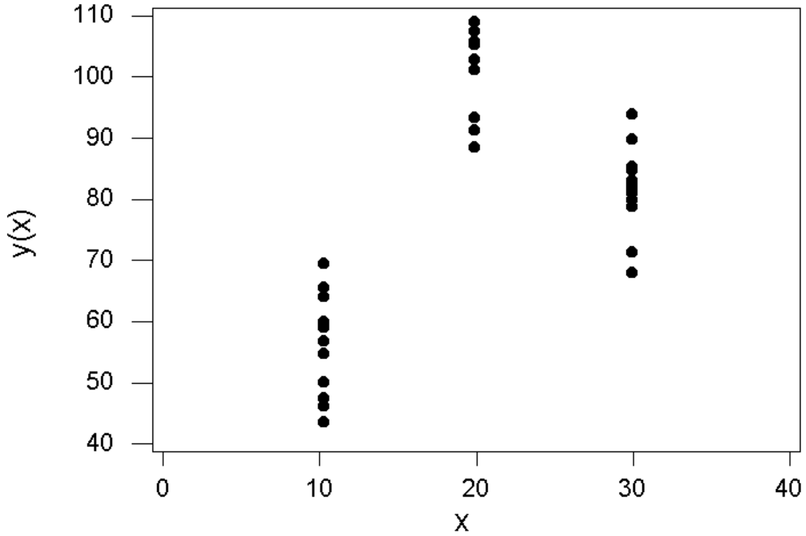
Chapter 11: Response Surface Experiments

What Function Can You Fit?



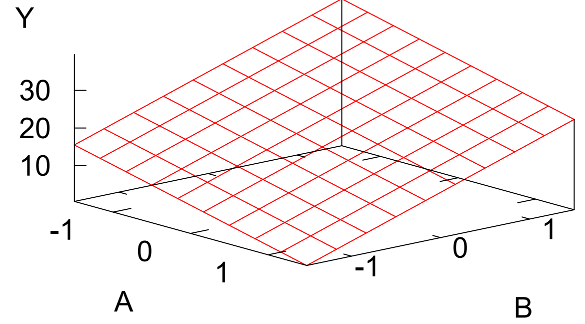
- With only two levels of x, a simple linear model is all we can fit.
- r^2 might be high, but what does it mean?

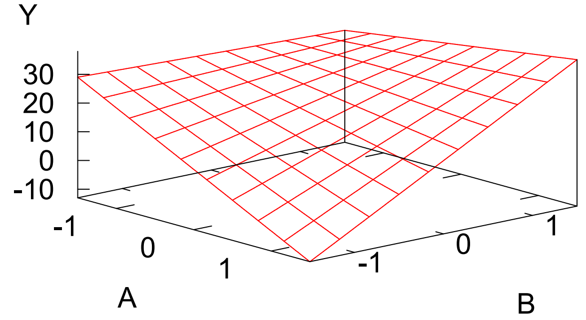
What Function Can You Fit?

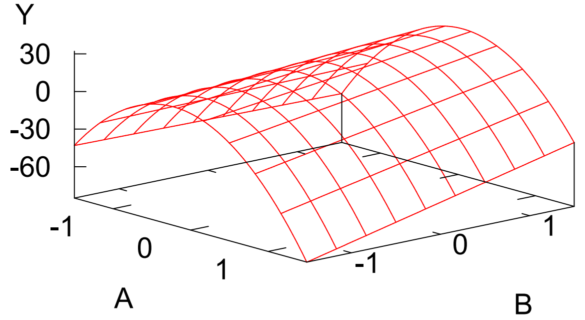


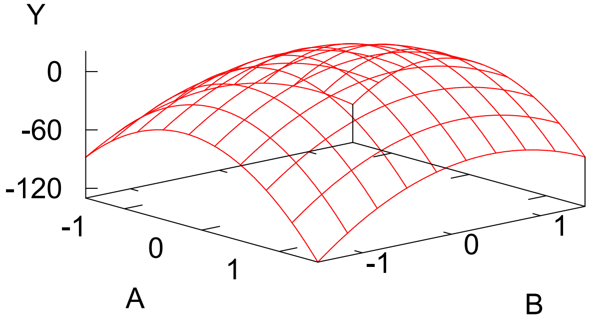
- At least three levels are necessary to detect lack of linear fit.
- r^2 and lack-of-fit are different issues. r^2 is not always a good lack-of-fit detector.
- The meaning of r^2 is limited to the data being analyzed.
- Our goal is to fit models that can resolve quadratic terms:

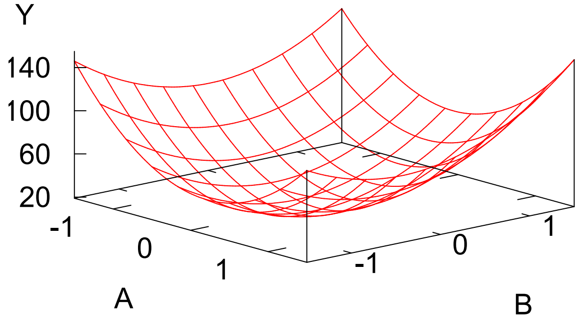
$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_{12} x_{12} + \cdots + b_{11} x_1^2 + b_{22} x_2^2 + \cdots$$

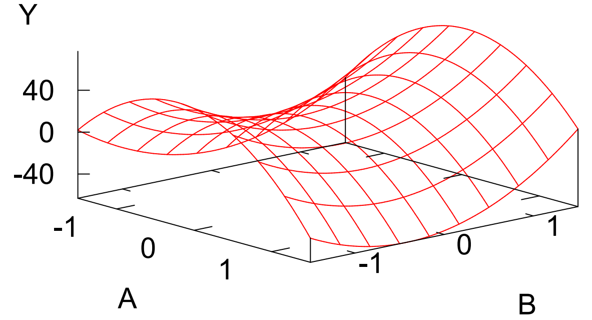

A: Y(A,B)=20-5A+8B


B: Y(A,B)=20-5A+8B+6AB


C: Y(A,B)=20-5A+8B+6AB-32A^2


D: Y(A,B)=20-5A+8B+6AB-32A^2-20B^2


E: Y(A,B)=20-5A+8B+6AB+32A^2+20B^2


F: Y(A,B)=20-5A+8B+6AB-32A^2+20B^2

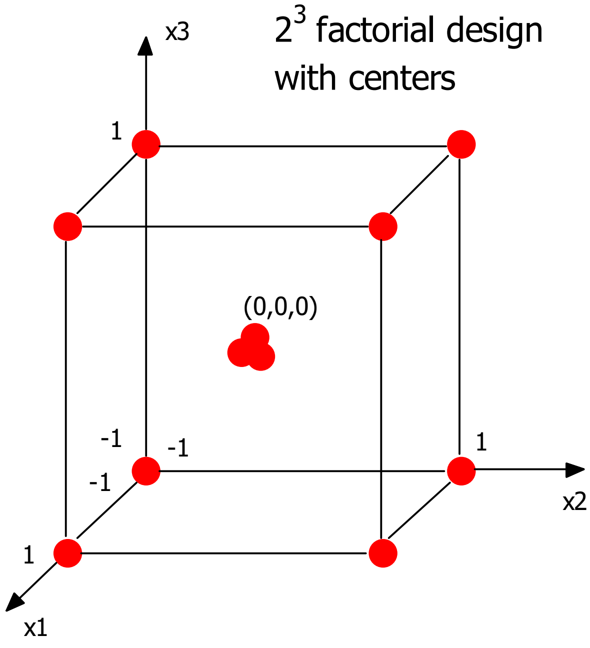
Response Surface Designs

- To use a response surface design:
  - All design variables must be quantitative!
  - Must have three or more levels of each variable.
- Available designs:
  - 2^k plus centers designs
    - Not true response surface designs.
    - Can detect the presence of curvature but can’t determine its source.
  - 3^k designs
  - Box-Behnken designs - $BB(k)$
  - Central composite designs - $CC(2^k)$

2^k Plus Centers Designs

- Consider the 2^3 plus centers design:

```
Row  x1  x2  x3  x12  x13  x23  x11  x22  x33
  1  -1  -1  -1    1    1    1    1    1    1
  2  -1  -1   1    1   -1   -1    1    1    1
  3  -1   1  -1   -1    1   -1    1    1    1
  4  -1   1   1   -1   -1    1    1    1    1
  5   1  -1  -1   -1   -1    1    1    1    1
  6   1  -1   1   -1    1   -1    1    1    1
  7   1   1  -1    1   -1   -1    1    1    1
  8   1   1   1    1    1    1    1    1    1
  9   0   0   0    0    0    0    0    0    0
```



2^k Plus Centers Designs

- There are three levels of each variable but ...

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_{12}x_{12} + \cdots + b_{**}x_*^2$$

where

$$b_{11} + b_{22} + \ldots \rightarrow b_{**}$$

- b_{**} provides a lack of fit test but nothing more.
- What we really wanted is:

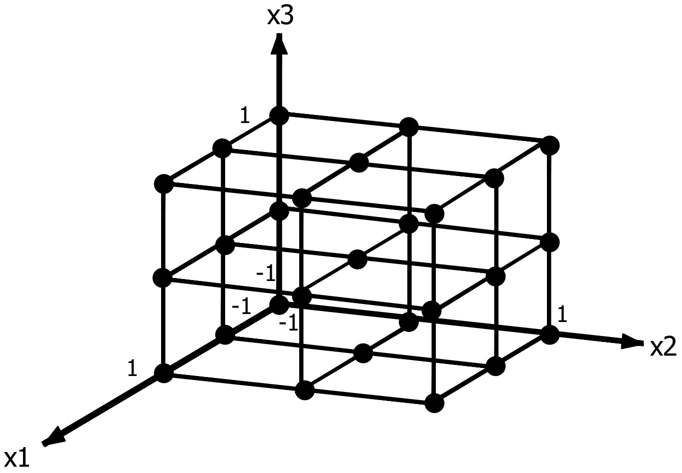
$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_{12}x_{12} + \cdots + b_{11}x_1^2 + b_{22}x_2^2 + \cdots$$

What designs can deliver this model?

The 3^k Factorial Designs

- Three levels of each of k quantitative variables.
- All possible combinations of levels: 3^k.
- Consider the 3^3 design:

```
Row  x1  x2  x3
  1  -1  -1  -1
  2  -1  -1   0
  3  -1  -1   1
  4  -1   0  -1
  5  -1   0   0
  6  -1   0   1
  7  -1   1  -1
  8  -1   1   0
  9  -1   1   1
 10   0  -1  -1
 11   0  -1   0
 12   0  -1   1
 13   0   0  -1
 14   0   0   0
 15   0   0   1
 16   0   1  -1
 17   0   1   0
 18   0   1   1
 19   1  -1  -1
 20   1  -1   0
 21   1  -1   1
 22   1   0  -1
 23   1   0   0
 24   1   0   1
 25   1   1  -1
 26   1   1   0
 27   1   1   1
```



The 3^k Factorial Designs

- The model will be:

$$\begin{aligned} y = {} & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_{12} + b_{13} x_{13} + b_{23} x_{23} \\ & + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 \end{aligned}$$

- The degrees of freedom:

$$\begin{aligned} Runs &= 3^3 = 27 \\ df_{total} &= 27 - 1 = 26 \\ df_{model} &= 3 + 3 + 3 = 9 \\ df_{\epsilon} &= 26 - 9 = 17 \end{aligned}$$

  and Occam will probably free up more error degrees of freedom.
- This is not an efficient use of resources.

$BB(3)$

x_1	x_2	x_3	*Runs*
±1	±1	0	4
±1	0	±1	4
0	±1	±1	4
0	0	0	3
Total Runs			15

$BB(4)$

Block	x_1	x_2	x_3	x_4	*Runs*
1	±1	±1	0	0	4
1	0	0	±1	±1	4
1	0	0	0	0	1
2	±1	0	0	±1	4
2	0	±1	±1	0	4
2	0	0	0	0	1
3	±1	0	±1	0	4
3	0	±1	0	±1	4
3	0	0	0	0	1
Total Runs					27

$BB(5)$

Block	x_1	x_2	x_3	x_4	x_5	*Runs*
1	±1	±1	0	0	0	4
1	0	0	±1	±1	0	4
1	0	±1	0	0	±1	4
1	±1	0	±1	0	0	4
1	0	0	0	±1	±1	4
1	0	0	0	0	0	3
2	0	±1	±1	0	0	4
2	±1	0	0	±1	0	4
2	0	0	±1	0	±1	4
2	±1	0	0	0	±1	4
2	0	±1	0	±1	0	4
2	0	0	0	0	0	3
Total Runs						46

$BB(6)$

x_1	x_2	x_3	x_4	x_5	x_6	*Runs*
±1	±1	0	±1	0	0	8
0	±1	±1	0	±1	0	8
0	0	±1	±1	0	±1	8
±1	0	0	±1	±1	0	8
0	±1	0	0	±1	±1	8
±1	0	±1	0	0	±1	8
0	0	0	0	0	0	6
Total Runs						54

$BB(7)$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	*Runs*
0	0	0	±1	±1	±1	0	8
±1	0	0	0	0	±1	±1	8
0	±1	0	0	±1	0	±1	8
±1	±1	0	±1	0	0	0	8
0	0	±1	±1	0	0	±1	8
±1	0	±1	0	±1	0	0	8
0	±1	±1	0	0	±1	0	8
0	0	0	0	0	0	0	6
Total Runs							62

$CC(2^2)$

x_1	x_2	*Runs*
±1	±1	4
0	0	5
±1.41	0	2
0	±1.41	2
Total Runs		13

$CC(2_V^{8-2})$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	*Runs*
±1	±1	±1	±1	±1	±1	1234	1256	64
0	0	0	0	0	0	0	0	10
±2.83	0	0	0	0	0	0	0	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	0	0	±2.83	2
Total Runs								90

$CC(2^3)$

x_1	x_2	x_3	*Runs*
±1	±1	±1	8
0	0	0	6
±1.68	0	0	2
0	±1.68	0	2
0	0	±1.68	2
Total Runs			20

$CC(2_{VII}^{7-1})$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	*Runs*
±1	±1	±1	±1	±1	±1	123456	64
0	0	0	0	0	0	0	14
±2.83	0	0	0	0	0	0	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	0	±2.83	2
Total Runs							92

$CC(2^4)$

x_1	x_2	x_3	x_4	*Runs*
±1	±1	±1	±1	16
0	0	0	0	7
±2	0	0	0	2
⋮	⋮	⋮	⋮	⋮
0	0	0	±2	2
Total Runs				31

$CC(2_{VI}^{6-1})$

x_1	x_2	x_3	x_4	x_5	x_6	*Runs*
±1	±1	±1	±1	±1	12345	32
0	0	0	0	0	0	9
±2.38	0	0	0	0	0	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	0	±2.38	2
Total Runs						53

$CC(2^5)$

x_1	x_2	x_3	x_4	x_5	*Runs*
±1	±1	±1	±1	±1	32
0	0	0	0	0	10
±2.38	0	0	0	0	2
⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	±2.38	2
Total Runs					52

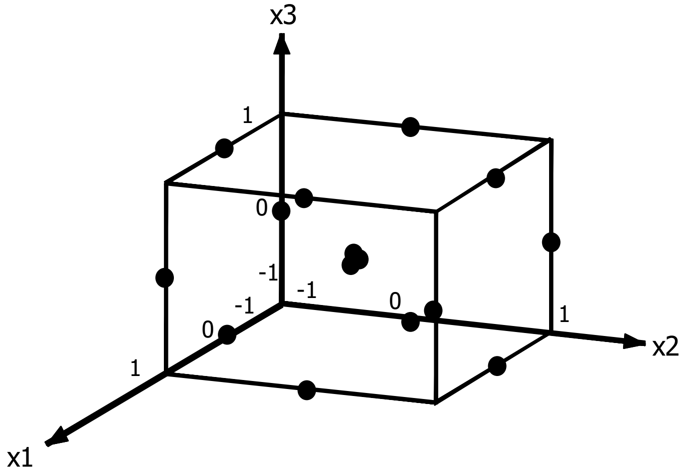
$CC(2_V^{5-1})$

x_1	x_2	x_3	x_4	x_5	*Runs*
±1	±1	±1	±1	1234	16
0	0	0	0	0	6
±2	0	0	0	0	2
⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	0	±2	2
Total Runs					32

The Box-Behnken Design

- Three levels of k variables.
- Kind of a fraction of the 3^k design with extra center cells.
- Consider the $BB(3)$ design:

```
Row   A   B   C
  1  -1  -1   0
  2   1  -1   0
  3  -1   1   0
  4   1   1   0
  5  -1   0  -1
  6   1   0  -1
  7  -1   0   1
  8   1   0   1
  9   0  -1  -1
 10   0   1  -1
 11   0  -1   1
 12   0   1   1
 13   0   0   0
 14   0   0   0
 15   0   0   0
```



The Box-Behnken Design

- The model will be:

$$\begin{aligned} y = {} & b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_{12} + b_{13} x_{13} + b_{23} x_{23} \\ & + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 \end{aligned}$$

- The degrees of freedom:

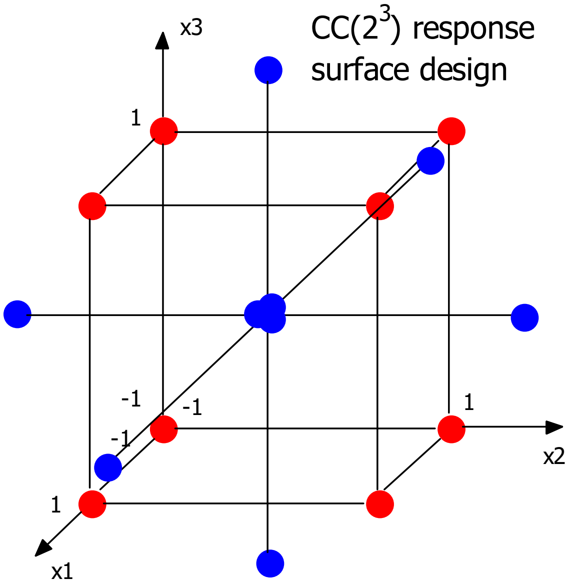
$$\begin{aligned} Runs & = 15 \\ df_{total} & = 15 - 1 = 14 \\ df_{model} & = 3 + 3 + 3 = 9 \\ df_{\epsilon} & = 14 - 9 = 5 \end{aligned}$$

  and Occam will probably free up more error degrees of freedom.
- Blocking is available.

The Central Composite Designs

- Based on the 2^k and 2^{k-p} designs.
- Center cells and star points added.
- Five levels of each variable.
- Consider the $CC(2^3)$ design:

```
Row  x1      x2      x3
  1  -1      -1      -1
  2  -1      -1       1
  3  -1       1      -1
  4  -1       1       1
  5   1      -1      -1
  6   1      -1       1
  7   1       1      -1
  8   1       1       1
  9   0       0       0

 10  -1.68    0       0
 11   1.68    0       0
 12   0      -1.68    0
 13   0       1.68    0
 14   0       0      -1.68
 15   0       0      -1.68
 16   0       0       0
 17   0       0       0
 18   0       0       0
 19   0       0       0
 20   0       0       0
```



The Central Composite Designs

- The model will be:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_{12} + b_{13} x_{13} + b_{23} x_{23} + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2$$

- The degrees of freedom:

$$\begin{aligned} Runs &= 8+6+6 = 20 \\ df_{total} &= 20-1 = 19 \\ df_{model} &= 3+3+3 = 9 \\ df_{\epsilon} &= 19-9 = 10 \end{aligned}$$

and Occam will probably free up more error degrees of freedom.

Comparison of the Five Variable Experiments

Design	Runs	df_{total}	df_{model}	df_{ϵ}
3^5	243	242	20	222
$BB(5)$	46	45	20	25
$CC(2_V^{5-1})$	32	31	20	11

and Occam will free up more error degrees of freedom.

Comparison of the Designs: Sample Size

- 3^k experiments are inefficient and don't get built.
- The sample size for $BB(3)$ is smaller than the sample size for $CC(2^3)$ so more $BB(3)$ experiments get built.
- The sample size for $CC(2_V^{5-1})$ is smaller than the sample size for $BB(5)$ so more $CC(2_V^{5-1})$ experiments get built.

Comparison of the Designs: Knowledge of the Design Space

- Different strategies are used for when you know and don't know the limitations of the variables.
- When you know safe limits for all of the design variables consider using the BB designs.
- When you don't know safe limits for all of the design variables consider using the CC designs.

Response Surface Designs in MINITAB

- Use **Stat**> **DOE**> **Response Surface**> **Create Response Surface Design** to create a design.
- Use **Stat**> **DOE**> **Response Surface**> **Define Custom Response Surface Design** to specify an existing design so that MINITAB will recognize it.
- Use **Stat**> **DOE**> **Response Surface**> **Analyze Response Suface Design** to analyze the data.
  - Enter the response in the **Responses**: window.
  - Specify the terms to be included in the model in the **Terms** window.
  - Turn on residuals diagnostic graphs in the **Graphs** window.
- Use **Stat**> **DOE**> **Response Surface**> **Contour/Surface Plots** to create multidimensional response surface plots.
- Use **Stat**> **DOE**> **Response Surface**> **Response Optimizer** to find the values of the design variables that will meet a specified response goal where the response can be a minimum, a maximum, or a target

Response Surface Designs in NCSS

Create a response surface experiment using **Analysis**> **Design of Experiments**> **Response Surface Designs**:

- Select the type of design in **Design Type**.
- Specify a column for the response in **Simulated Response Variable** (e.g. c1 or Y).
- Specify a column for blocks in **Block Variable** (e.g. c2 or Blocks).
- Specify the column for the first design variable in **First Factor Variable** (e.g. c3 or A).
- Specify the factor levels in **Factor Values**. The values -1 and $+1$ are recommended and 0 is assumed for the center level. Specify a set of levels for as many variables as are required for the design.
- Replicate the design manually with copy/paste operations and define each replicate as a new block.

Analyze the experiment using **Analysis**> **Design of Experiments**> **Analysis of Response Surface Designs** or **Analysis**> **Regression/Correlation**> **Multiple Regression** (**2001 Edition**). See the notes from Chapter 9 for details for configuring these analyses.

Putting It All Together

The following algorithm assumes that you're starting with a process that you have little to no experience with. If you do have some knowledge of the system, you may be able to start from a later step.

1. Follow the 11-step DOE process.
2. Use a fractional factorial or Plackett-Burman design to identify the vital few variables from the many variables.Run the fold-over design to identify significant two-factor interactions.
3. Run a 2^k or 2^{k-1} with centers design to quantify main effects, two factor interactions, and to test for curvature in the response space.
4. Run a response surface design, e.g. $BB(k)$ or $CC(2^{k-p})$, to quantify main effects, two factor interactions, and quadratic terms. Build the experiment in blocks if possible so that you can suspend the experiment if all of the answers are apparent early.
5. Build a confirmation experiment to test the results from new runs to the model prediction.

Strategies for Missing Runs and Outliers

- Missing runs from an otherwise good experiment design cause undesireable correlations between predictors.
- Outliers are unusual observations, hopefully with an obvious special cause, that deviate substantially from their predicted values.
- Outliers should never be removed without cause. When there is sufficient cause, an outlier should be replaced with a new observation or can be treated like a missing value.
- Determine if the missing runs and outliers are *missing with cause (MWC)* or *missing at random (MAR)*.
- If observations are missing with cause, search the cause out and take appropriate action. For example, if observations are missing because one level of a design variable was chosen poorly, remove all of the observations made at that level and analyze what's left.
- If the observations are missing at random, then the analysis can be corrected to account for them using the *imputation* procedure below.
- If possible, for observations missing at random, build replacement runs to fill in the missing values. Consider building some of the runs that survived (center point runs are a good choice) with those to confirm that the process hasn't shifted between the original and replacement runs.
- If the design is replicated, df_ϵ is very large, and the number of missing values is relatively small compared to df_ϵ, replace the missing observations with the average of their cell means and complete the regular analysis.
- To impute observations missing at random, treat the missing values as predictors in the model by simultaneous solution of the least square system of equations:

  $$\frac{\partial}{\partial \hat{y}_i} \sum \epsilon_i^2 = 0$$

  or, find the optimal $\hat{y}_i$ values by:

  1. Replace the missing values with best guesses, such as the grand or cell means
  2. Fit the desired model and store the predicted values
  3. Replace the initial guesses with predicted values
  4. Repeat steps 2 and 3 until the predicted values converge (note: convergence corresponds to $\epsilon_i = 0$).
  5. If the number of missing values is substantial compared to the ANOVA's df_ϵ, reduce df_ϵ by the number of missing observations and recalculate the ANOVA table and regression coefficient standard errors, t values, and p values.
- Always be absolutely clear about how you handled the missing values in reporting any results.