MOST PEOPLE WOULD RATHER LIVE WITH A PROBLEM THEY CAN'T SOLVE, THAN ACCEPT A SOLUTION THEY CAN'T UNDERSTAND.

R. D. WOOLSEY / H. S. SWANSON

Analyze

Analyze is presented in the following topic areas:

- Measuring and modeling relationships between variables
- Hypothesis testing
- Failure mode and effects analysis (FMEA)
- Additional analysis methods

Measuring and Modeling Relationships Between Variables

Measuring and modeling relationships between variables is reviewed in the following topic areas:

- Regression
- Correlation coefficient
- Multivariate tools

Note that the authors have presented regression ahead of the correlation coefficient for explanation purposes.

Regression

Simple linear regression and multiple linear regression will be discussed here. Note that non-linear regression models will not be tested on the CSSBB exam, and will not be described here. The material on linear regression may be found in several statistics books, including Triola (1994)[24].

Simple Linear Regression Model

Consider the problem of predicting the test results (y) for students based upon an input variable (x), the amount of preparation time in hours using the data presented in Table 8.1. Please note that this is hypothetical data, and is not based on actual results.

Simple Linear Regression Model (Continued)

Student	Study Time (Hours)	Test Results (%)
1	60 h	67%
2	40 h	61%
3	50 h	73%
4	65 h	80%
5	35 h	60%
6	40 h	55%
7	50 h	62%
8	30 h	50%
9	45 h	61%
10	55 h	70%

Table 8.1 Study Time Versus Test Results

An initial approach to the analysis of the data in Table 8.1 is to plot the points on a graph known as a scatter diagram, as shown in Figure 8.2. Observe that y appears to increase as x increases. One method of obtaining a prediction equation relating y to x is to place a ruler on the graph and move it about until it seems to pass through the majority of the points, thus providing what is regarded as the "best fit" line.

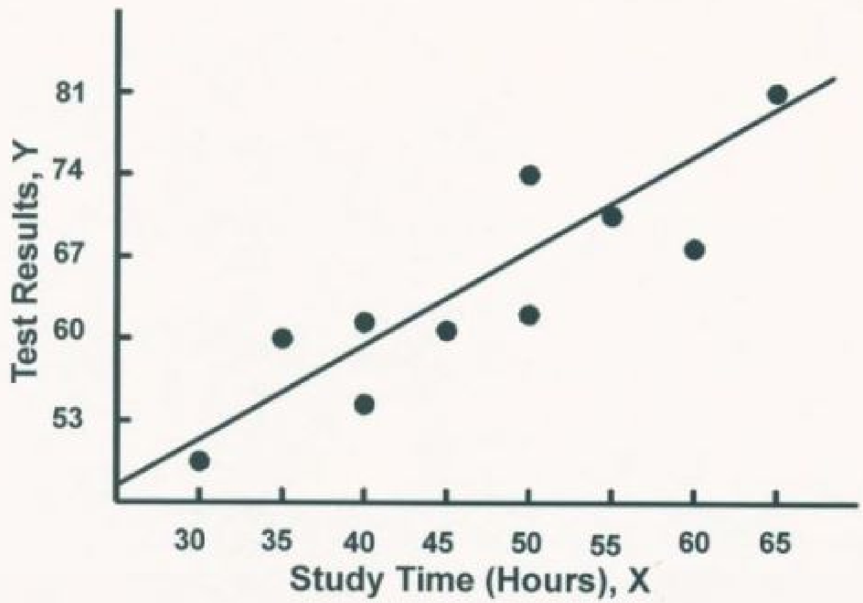


Figure 8.2 Plot of Study Time Versus Test Results

Simple Linear Regression Model (Continued)

The mathematical equation of a straight line is:

$$y = \beta_0 + \beta_1 x$$

Where β_0 is the y intercept when x = 0 and β_1 is the slope of the line. Please note in Figure 8.2 that the x axis does not go to zero so the y intercept appears too high. The equation for a straight line in this example is too simplistic. There will actually be a random error which is the difference between an observed value of y and the mean value of y for a given value of x. One assumes that for any given value of x, the observed value of y varies in a random manner and possesses a normal probability distribution. The concept is illustrated in Figure 8.3:

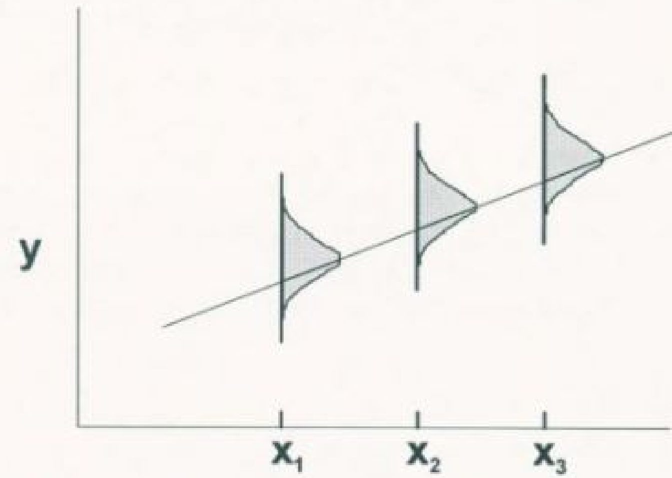


Figure 8.3 Variation in y as a Function of x

The probabilistic model for any particular observed value of y is:

$$y = \begin{pmatrix} \text{Mean value of y for} \\ \text{a given value of x} \end{pmatrix} + (\text{random error})$$

$$y = \beta_0 + \beta_1 x + \varepsilon$$

The Method of Least Squares

The statistical procedure of finding the "best-fitting" straight line is, in many respects, a formalization of the procedure used when one fits a line by eye. The objective is to minimize the deviations of the points from the prospective line.

If one denotes the predicted value of y obtained from the fitted line as $\hat{y}$ the prediction equation is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Where: $\hat{\beta}_0$ and $\hat{\beta}_1$ represent estimates of the true β_0 and β_1, as shown in Figure 8.4.

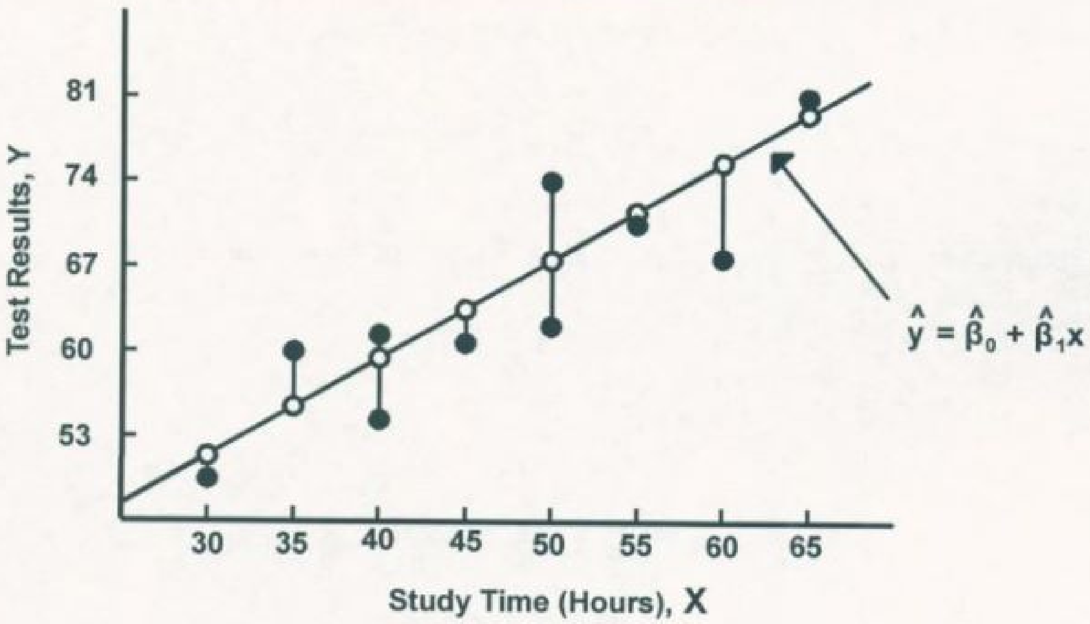


Figure 8.4 Study Time Versus Test Results

Having decided to minimize the deviation of the points in choosing the best fitting line, one must now define what is meant by "best."

The Method of Least Squares (Continued)

The best fit criterion of goodness known as the principle of least squares is employed:

> Choose, as the best fitting line, the line that minimizes the sum of squares of the deviations of the observed values of y from those predicted.

Expressed mathematically, minimize the sum of squared errors given by:

$$SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

Substituting for $\hat{y}_i$ one obtains the following expression:

$$\text{Sum of squared errors} = SSE = \sum_{i=1}^{n}\left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]^2$$

The least square estimator of β_0 and β_1 are calculated as follows:

$$S_{x^2} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{x^2}} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Once $\hat{\beta}_0$ and $\hat{\beta}_1$ have been computed, substitute their values into the equation of a line to obtain the least squares prediction equation, or regression line.

As noted earlier, the prediction equation for $\hat{y}$ is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Where: $\hat{\beta}_0$ and $\hat{\beta}_1$ represent estimates of the true β_0 and β_1.

Least Squares Example

Example 8.1: Obtain the least squares prediction line for the data below:

I	x_i	y_i	x_i^2	x_iy_i	y_i^2
1	60	67	3,600	4,020	4,489
2	40	61	1,600	2,440	3,721
3	50	73	2,500	3,650	5,329
4	65	80	4,225	5,200	6,400
5	35	60	1,225	2,100	3,600
6	40	55	1,600	2,200	3,025
7	50	62	2,500	3,100	3,844
8	30	50	900	1,500	2,500
9	45	61	2,025	2,745	3,721
10	55	70	3,025	3,850	4,900
Sum	470	639	23,200	30,805	41,529

Table 8.5 Data for the Study Time versus Test Results Example

$$S_{x^2} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 23{,}200 - \frac{(470)^2}{10} = 1{,}110$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} = 30{,}805 - \frac{(470)(639)}{10} = 772$$

$$\bar{x} = \frac{470}{10} = 47 \qquad \bar{y} = \frac{639}{10} = 63.9$$

Least Squares Example (Continued)

Example 8.1 (continued):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{x^2}} = \frac{772}{1,110} = 0.6955$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 63.9 - (0.6955)(47) = 31.2115$$

$$\hat{y} = 31.2115 + 0.6955\,x$$

One may predict y for a given value of x by substitution into the prediction equation. For example, if 60 hours of study time is allocated, the predicted test score would be:

$$\hat{y} = 31.2115 + (0.6955)(60)$$

$$\hat{y} = 72.9415 = 73\%$$

Hints on Regression Analysis

- Be careful of rounding errors. Normally, the calculations should carry a minimum of six significant figures in computing sums of squares of deviations. Note that the prior example consisted of convenient whole numbers which does not occur often.
- Always plot the data points and graph the least squares line. If the line does not provide a reasonable fit to the data points, there may be a calculation error.
- Projecting a regression line outside of the test area can be risky. The above equation suggests, without study, a student would make 31% on the test. The odds favor 25% if answer a is selected for all questions. The equation also suggests that with 100 hours of study the student should attain 100% on the examination - which is highly unlikely.

Calculating s_e^2, an Estimator of σ_ε^2

Recall, the model for y assumes that y is related to x by the equation:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

If the least squares line is used:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

A random error, ε, enters into the calculations of β_0 and β_1. The random errors affect the error of prediction. Consequently, the variability of the random errors (measured by σ_ε^2) plays an important role when predicting by the least squares line.

The first step toward acquiring a boundary on a prediction error requires that one estimates σ_ε^2. It is reasonable to use SSE (sum of squares for error) based on (n - 2) degrees of freedom, one for each variable (x and y).

An Estimator for σ_ε^2

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n-2}$$

$\hat{\sigma}_\varepsilon^2$ is sometimes shown as s_e^2

Formula for Calculating SSE

Sum of squared errors = SSE

$$SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]^2$$

SSE may also be written:

$$SSE = S_{y^2} - \hat{\beta}_1 S_{xy} = S_{y^2} - \frac{\left(S_{xy}\right)^2}{S_{x^2}}$$

Where:

$$S_{y^2} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

S_{xy} and S_{x^2} were previously defined.

Calculating s^2 (Continued)

Example 8.2: Calculate an estimated σ_ε^2 for the data in Table 8.5.

$$S_{y^2} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 41{,}529 - \frac{(639)^2}{10} = 696.9$$

$$SSE = S_{y^2} - \hat{\beta}_1 S_{xy} = 696.9 - (0.6955)(772) = 159.97$$

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n-2} = \frac{159.97}{10-2} = 19.996$$

$$\hat{\sigma}_\varepsilon = 4.47$$

How can one interpret the values of SSE and $\hat{\sigma}_\varepsilon^2$? Refer to Figure 8.4 and note the deviations of the 10 points from the least squares line. The sum SSE = 159.97 is equal to the sum of squares of the numerical values of these deviations.

$\hat{\sigma}_\varepsilon$ from the above calculation equals 4.47. Thus, most of the points will fall within $\pm 1.96\hat{\sigma}_\varepsilon$ or 8.76 of the line. Approximately 95% of the values should be in this region. In Figure 8.4, all of the values are within ± 8.76 of the line. This estimate provides a rough check on the calculated value of $\hat{\sigma}_\varepsilon$.

Inferences Concerning the Slope β_1 of a Line

The existence of a significant relationship between y and x can be tested by whether β_1 is equal to 0. If $\beta_1 \neq 0$ there is a linear relationship. The null hypothesis and alternative hypothesis are:

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0$$

The test statistic is a t distribution with n - 2 degrees of freedom:

$$t = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \qquad \text{where:} \quad s_{\hat{\beta}_1} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}}$$

Inferences Concerning the Slope β_1 of a Line (Continued)

Example 8.3: From the data in Table 8.5, determine if the slope results are significant at a 95% confidence level.

$$t = \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} = \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}}\right)} = \frac{0.6955 - 0}{\left(\frac{4.47}{\sqrt{1,110}}\right)} = 5.18$$

For a 95% confidence level, determine the critical values of t with $\alpha = 0.025$ in each tail, using n - 2 = 8 degrees of freedom: $t_{0.025,8} = -2.306$ and $t_{0.025,8} = 2.306$. Reject the null hypothesis if $t > 2.306$ or $t < -2.306$, depending on whether the slope is positive or negative. In this case, the null hypothesis is rejected and we conclude that $\beta_1 \neq 0$ and there is a linear relationship between y and x.

Confidence Interval Estimate for the Slope β_1

The confidence interval estimate for the slope β_1 is given by:

$$\hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}} \quad \text{thus,}$$

$$\hat{\beta}_1 - t_{\alpha/2,\, n-2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2,\, n-2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}}$$

Example 8.4: Substitute previous data into the above formula to obtain the confidence interval around the slope of the line.

$$\hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{S_{x^2}}}$$

$$0.6955 \pm 2.306 \frac{4.47}{\sqrt{1,110}}$$

$$0.3861 < \beta_1 < 1.0049$$

Intervals constructed by this procedure will enclose the true value of β_1 95% of the time. Hence, for every 10 hours of increased study, the expected increase in test scores would fall in the interval of 3.86 to 10.05 percentage points.

Multiple Linear Regression

Multiple linear regression is an extension of the methodology for linear regression to more than one independent variable. By including more than one independent variable, a higher proportion of the variation in y may be explained.

A full explanation of multiple linear regression is left to the resources of the readers of this Primer. A few models and definitions are presented:

A First-Order Linear Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

A Second-Order Linear Model (Two Predictor Variables)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon$$

$$s^2 = \frac{SSE}{n - (k + 1)}$$

Just like r^2 (the linear coefficient of determination) R^2 (the multiple coefficient of determination) take values in the interval:

$$0 \leq R^2 \leq 1$$

Source	DF	SS	MS
Regression Error	k n-(k+1)	SSR SSE	MSR=SSR/k MSE=SSE/[–(k+1)]
Total	n-1	Total SS	

Table 8.6 ANOVA Table for Multiple Regression Analysis

Note: k = the number of predictor variables

Correlation Coefficient

The population linear correlation coefficient, ρ, measures the strength of the linear relationship between the paired x and y values in a population. ρ is a population parameter. For the population, the Pearson product moment coefficient of correlation, $\rho_{x,y}$ is given by:

$$\rho_{x,y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y}$$

Where cov means covariance. Note that $-1 \leq \rho \leq +1$

The sample linear correlation coefficient, r, measures the strength of the linear relationship between the paired x and y values in a sample. r is a sample statistic. For a sample, the Pearson product moment coefficient of correlation, $r_{x,y}$ is given by:

$$\boxed{r_{x,y} = \frac{S_{xy}}{\sqrt{S_{x^2} S_{y^2}}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Note that $-1 \leq r \leq +1$

Example 8.5: Using the study time and test score data reviewed earlier, determine the correlation coefficient.

Given:

$$S_{xy} = 772 \quad S_{x^2} = 1{,}110 \quad S_{y^2} = 696.9$$

Solution:

$$r_{x,y} = \frac{S_{xy}}{\sqrt{S_{x^2} S_{y^2}}} = \frac{772}{\sqrt{(1{,}110)(696.9)}} = 0.878$$

Note that the numerator used in calculating r is identical to the numerator of the formula for the slope β_1. Thus, the coefficient of correlation r will assume exactly the same sign as β_1 and will equal zero when $\beta_1 = 0$.

Value of r

- A positive value for r implies that the line slopes upward to the right.
- A negative value for r implies that the line slopes downward to the right.
- Note that r = 0 implies no linear correlation, not simply "no correlation." A pronounced curvilinear pattern may exist.

The material on linear correlation may be found in several statistics books, including Triola (1994)[24].

Correlation Coefficient (Continued)

When $r = 1$ or $r = -1$, all points fall on a straight line; when $r = 0$, they are scattered and give no evidence of a linear relationship. Any other value of r suggests the degree to which the points tend to be linearly related.

If x is of any value in predicting y, then SSE can never be larger than:

$$S_{y^2} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Because:

$$SSE = S_{y^2} - \hat{\beta}_1 S_{xy} = S_{y^2} - \frac{(S_{xy})^2}{S_{x^2}}$$

Coefficient of Determination (R^2)

The coefficient of determination is R^2. The square of the linear correlation coefficient is r^2. It can be shown that: $R^2 = r^2$

$$R^2 = r^2 = \frac{S_{y^2} - SSE}{S_{y^2}} = 1 - \frac{SSE}{S_{y^2}} = \frac{(S_{xy})^2}{S_{x^2}S_{y^2}}$$

The coefficient of determination is the proportion of the explained variation divided by the total variation, when a linear regression is performed. r^2 lies in the interval of $0 \leq r^2 \leq 1$. r^2 will equal +1 only when all the points fall exactly on the fitted line. That is, when SSE equals zero.

Example 8.6: Using the data from Example 8.5, determine the coefficient of determination.

$$r^2 = \frac{(S_{xy})^2}{S_{x^2}S_{y^2}} = \frac{(772)^2}{(1{,}110)(696.9)} = 0.771$$

$$\text{or} \quad r^2 = (0.878)^2 = 0.771$$

One can say that 77% of the variation in test scores can be explained by variation in study hours.

Correlation Example

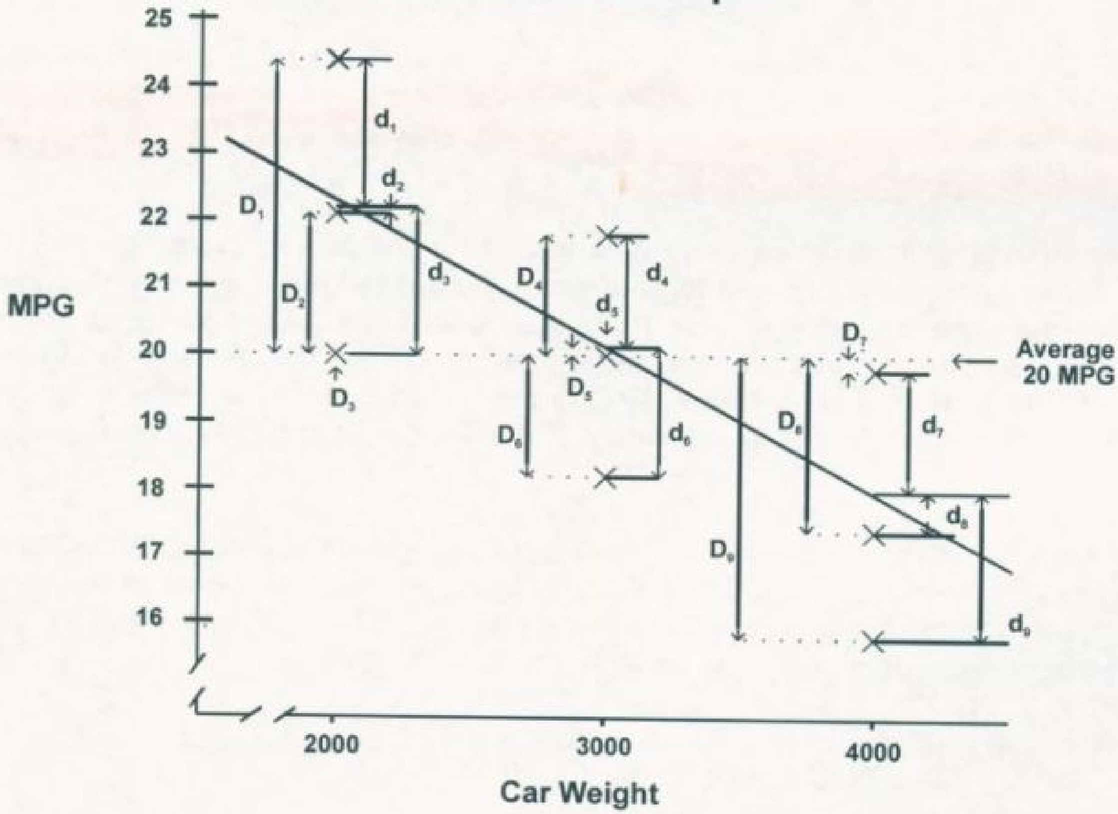


Figure 8.7 Correlation Plot of Car Weight and MPG

$$SST = \sum_{i=1}^{n} D_i^2 = D_1^2 + D_2^2 + \ldots + D_9^2 \text{ and } SSE = \sum_{i=1}^{n} d_i^2 = d_1^2 + d_2^2 + \ldots + d_9^2$$

$$r^2 = 1 - \frac{SSE}{SST} \text{ or } \frac{SST - SSE}{SST}$$

Where SST = total sum of squares (from the experimental average) and SSE = total sum of squared errors (from the best fit). Note that when SSE is zero, r^2 equals one and when SSE equals SST, then r^2 equals zero.

Correlation Versus Causation

In the above example, there is strong evidence of a correlation between car weight and gas milage. The student should be aware that a number of other factors (carburetor type, car design, air conditioning, passenger weights, speed, etc.) could also be important. The most important cause may be a different or a collinear variable. For example, car and passenger weight may be collinear. There can also be such a thing as a nonsensical correlation, i.e. it rains after my car is washed.

Multivariate Tools

In univariate statistics there are one or more independent variables (X1, X2), and only one dependent variable (Y). Multivariate analysis is concerned with two or more dependent variables, Y1, Y2, being simultaneously considered for multiple independent variables, X1, X2, etc.

The manual effort used to solve multivariate problems was an obstacle to its earlier use. Recent advances in computer software and hardware have made it possible to solve more problems using multivariate analysis. Some of the software programs available to solve multivariate problems include: SPSS, S-Plus, SAS, and Minitab. This coverage of multivariate analysis can only be considered an introduction to the subject. For more in-depth information the reader is advised to consult other references.

Multivariate analysis has found wide usage in the social sciences, psychology, and educational fields. Applications for multivariate analysis can also be found in the engineering, technology, and scientific disciplines. This element will highlight the following multivariate concepts or techniques:

- Factor analysis
- Discriminant function analysis
- Cluster analysis
- Canonical correlation analysis
- Multivariate analysis of variance

Factor Analysis

Factor analysis is a data reduction technique to identify factors that explain variation. It is very similar to the principal components analysis technique. That is, factor analysis attempts to simplify complex sets of data, reducing many factors to a smaller set. However, there is some subjective judgment involved in describing the factors in this method of analysis. The output variables are linearly related to the input factors.

Factor Analysis (Continued)

The variables under investigation should be measurable, have a range of measurements, and be symmetrically distributed. There should be four or more input factors for each dependent variable. Factor analysis undergoes two stages: factor extraction and factor rotation. The first analysis in this discussion will distinguish the major factors for further study (extraction). The second stage will rotate the factors, to make them more meaningful.

A principal components analysis can be performed on some community evaluation data to provide a reduction in the number of factors. (Note: Minitab can also examine the data through a "maximum likelihood" method.) Economic development data could indicate that two factors are significant. From this information, a researcher can go into Minitab, and perform a factor analysis for these two factors and obtain a correlation matrix. See Table 8.8.

Unrotated Factor Loadings and Communalities			
Variable	Factor 1	Factor 2	Communality
creative class	-0.428	0.903	0.998
entrepreneurial	-0.961	-0.177	0.954
university-industry	-0.969	-0.075	0.945
high tech workers	-0.850	-0.144	0.743
venture capital	-0.905	-0.023	0.820
Variance	3.5856	0.8731	4.4587
% Var	0.717	0.175	0.892

Table 8.8. Principal Component Factor Analysis of the Correlation Matrix

To make sense of the information in Table 8.8, note that Factor 1 has the four factors in a grouping (entrepre, universi, high tech, and venture) and Factor 2 has the creative class as the prime factor. This is a similar result to the earlier principal components analysis. Again, the first factor has negative readings, so the researcher should examine that grouping more closely for meaning.

Factor Analysis (Continued)

The communality column indicates whether the chosen variables explain the variability fit very well. The communality numbers are very high. This means that the researcher can state that the two major factors in high technology community development would involve the five studied variables.

The data and factors can be rotated (by the software) to view the data from a different perspective. The four rotational methods in Minitab are: equimax, varimax, quartimax, and orthomax. Other software has other varieties. The student should consult Minitab or other references for additional rotational detail if desired.

Discriminant Analysis

If one has a sample with known groups, discriminant analysis can be used to classify the observations or attributes into two or more groups. Discriminant analysis can be used as either a predictive or a descriptive tool. The decisions could involve medical care, college success attributes, car loan credit worthiness, or the previous economic development issues. Discriminant analysis can be used as a follow-up to the use of MANOVA (the last area of coverage in this Section element). Again, linear combinations of predictors or groups are provided by the researcher.

The possible number of linear combinations (discriminant functions) for a study would be the smaller of the number of groups -1, or the number of variables. Some assumptions in discriminant analysis are: the variables are multivariately normally distributed, the population variances and covariances among the dependent variables are the same, and the samples within the variables are randomly obtained and exhibit independence of scores from the other samples.

Minitab provides two forms of analysis: linear and quadratic discriminant analysis. The linear discriminant analysis assumes that all groups have the same covariance matrix. This is not the case for the quadratic case. In linear discriminant analysis the Mahalanobis distance is the measure used to form or classify groups. The Mahalanobis distance is the squared distance (linear measure) from the observation to the group center. The classification into groups is formed by the distance measure.

In the quadratic discriminant analysis, the squared distance does not translate to a linear function, but into a quadratic function. The quadratic distance is called the generalized squared distance.

Discriminant Analysis (Continued)

Using provided information on high technology growth, a discriminant analysis is shown in Table 8.9. An additional column is used to state that the area is a "new economy" community. For example, a "yes" or "no" will used to indicate if a community is considered a "new economy" area. The discriminant analysis will correlate the data and verify if the decision was correct. (Note: The authors have not verified the information in any of the provided tables).

Factors						Dependent Response
community	creative class	entrepr-eneurial culture	university-industry projects	high tech workers	venture capital	new economy
Austin	15	600	500	100	4,000	yes
San Jose	25	1,000	1,000	500	6,000	yes
Chapel Hill	20	500	700	400	1,500	yes
San Diego	15	500	600	400	2,000	yes
Chicago	12	300	300	200	1,000	no
Minnea-polis	12	300	200	350	2,000	yes
Fargo	30	40	30	25	400	no
Terre Haute	10	50	50	10	300	no
St Louis	10	100	100	30	500	no
Seattle	25	300	500	300	3,000	yes

Table 8.9. High Technology Factors and the New Economy

The Minitab analysis of Table 8.9 is provided in Table 8.10 that follows. It states that the decisions on the grouping were 10 out of 10 (100% correct). That is, the values in the various factors match up enough to place various regions in certain categories.

Discriminant Analysis (Continued)

Discriminant Analysis: New economy versus creative class, entrepreneurial culture University-industry projects and venture capital.

Linear Method for Response: New Economy
Predictors: creative, entrepre, universi, high tech venture

Summary of Classification:

Put into true group?		
Group	no	yes
No	4	0
Yes	0	6
Total N	4	6
N Correct	4	6
Proportion	1.000	1.000

N = 10 N Correct = 10 Proportion Correct = 1.000 (100%)

Squared distance between groups
(also called the Mahalanobis distance):

	no	yes
no	0.00000	9.17913
yes	9.17913	0.00000

Linear Discriminant Function for Group:

	no	yes
constant	-3.2185	-7.5646
creative	0.4064	0.2751
entrepre	0.0221	0.0041
universi	-0.0141	-0.0054
high tech	0.0016	0.0225
venture	-0.0018	0.0010

Table 8.10. Discriminant Analysis Results of the New Economy

The above results are Minitab outputs with few adjustments. The student with access to Minitab might choose to change the classification of Seattle and Austin from new economy to not new economy to observe the interesting outcome.

Cluster Analysis

Cluster analysis is used to determine groupings or classifications for a set of data. A variety of rules or algorithms have been developed to assist in group formations. The natural groupings should have observations classified so that similar types are placed together. A file on attributes of high achieving students could be grouped or classified by IQ, parental support, school system, study habits, and available resources. Cluster analysis is used as a data reduction method in an attempt to make sense of large amounts of data from surveys, questionnaires, polls, test questions, scores, etc.

The economic development example in the previous discussion will again be used to validate groupings. The two types of groups will be new economy and not new economy. A graphic output from the analysis is the classification tree or dendogram. It is a graphic line graph linking variables and groups at various stages.

Cluster Analysis Example

The Table 8.9 data will be analyzed by the cluster analysis method. Using Minitab, the first analysis request calls for two groups. (More groups can be used.) It is displayed below in Table 8.11.

Cluster	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
1	9	13805274.000	1081.806	2393.876
2	1	0.000	0.000	0.000

Cluster Centroids:

Variable	Cluster1	Cluster2	Grand centroid
creative	16.5556	25.0000	17.4000
entrepre	298.8889	1000.0000	369.0000
universi	331.1111	1000.0000	398.0000
high tech	201.6667	500.0000	231.5000
venture	1633.3333	6000.0000	2070.0000

Table 8.11. Cluster Analysis of High Tech Communities. (Minitab)

Cluster Analysis Example (Continued)

The analysis shows our requested two groupings. However, instead of grouping into our presumed two groups of new economy or not new economy, the program used an algorithm based on measures of "closeness" between groups. Since the author requested two groups, the final iteration provides two groups. The dendogram in Figure 8.12 provides a visual that San Jose is distinctive and of a higher ranking than the other communities.

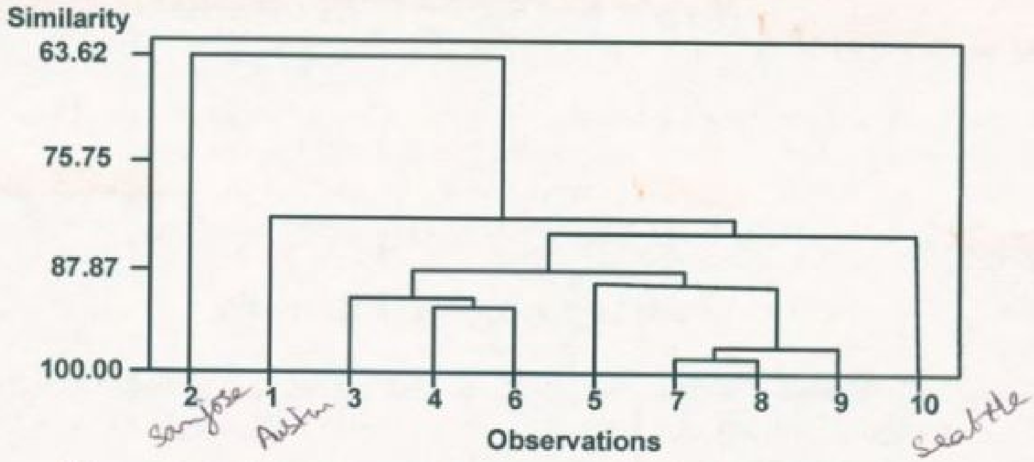


Figure 8.12. Dendogram of the High Tech Communities

The dendogram shows that Austin and Seattle are also distinct from the other lower communities. Communities 7, 8, and 9 form the lowest cluster. The student can verify this result by rerunning the analysis and requesting four groupings. Another interesting analysis would be to group the data by the original five factors as shown in Figure 8.13.

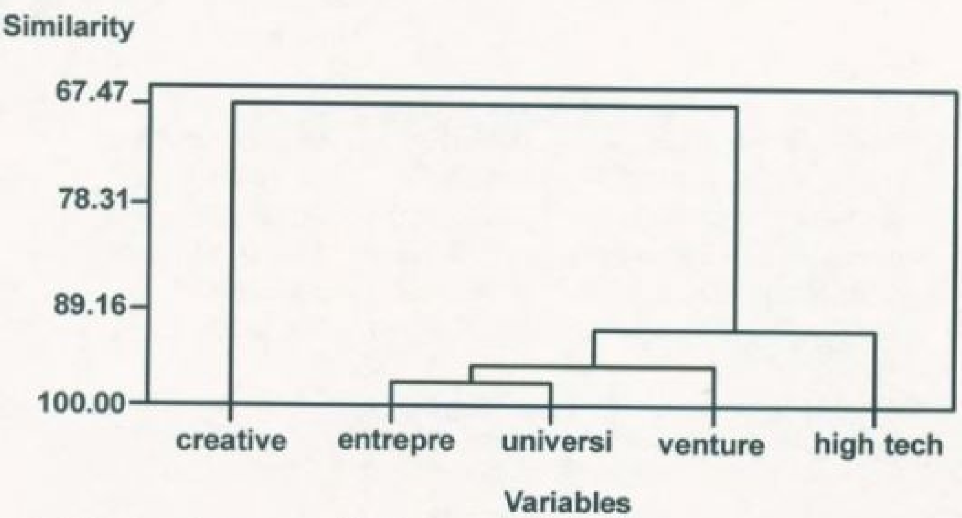


Figure 8.13. Dendogram of High Tech Communities by Factors (Minitab)

Cluster Analysis Example (Continued)

Figure 8.13 indicates that creative class is separated from the other factors in the grouping. In the principal components and factor analysis discussion, creative class was always the major factor listed in the second group, separated from the other four factors. Similar results that can be obtained using different multivariate tools.

Canonical Correlation Analysis

Canonical analysis tests the hypothesis that effects can have multiple causes and causes can have multiple effects. This technique was developed by Hotelling in 1935, but was not widely used for over 50 years. The emergence of personal computers and statistical software has led to its fairly recent adoption.

Canonical correlation analysis is a form of multiple regression to find the correlation between two sets of linear combinations. Each set may contain several related variables. The relating of one set of independent variables to one set of dependent variables will form linear combinations. The largest correlation values for sets are used in the analysis. The pairings of linear combinations are called canonical variates, and the correlations are called canonical correlations (also called characteristic roots). There may be more than one pair of linear combinations that could be applicable for an investigation. The maximum number of linear combinations would be limited by the number of variables in the smaller set. Most researchers involve only two sets.

The canonical correlation coefficient, r_c, is similar to the Pearson product-moment correlation coefficient. The rule of thumb is to have values above 0.30. The squared value would represent less than 10% in overlapping variance between pairs of canonical variates.

The linear combinations can be determined from linear matrix algebra or statistical software. For instance, SPSS software can test for significance of canonical correlation and will provide several additional tests. (Green, 1997)[7]

Table 8.14 illustrates the correlation of sets of independent variables to sets of dependent variables. An industrial survey can be conducted to see if there is a correlation between the characteristics of a quality engineer to the listed job skills of a quality engineer. There may be a set of variables that are strongly correlated and canonical correlation can be used.

Canonical Correlation Analysis (Continued)

Quality Engineer Characteristics	Job Skills Listed
Hands-on tendencies	ISO/TS 16949:2002 knowledge
On the floor activities	Autocad experience
Good problem solver	Automotive industry experience
Works well with people	Black belt certificate
Has sense of urgency	Good SPC skills

Table 8.14. Correlation Sets of Independent Variables

Hotelling's T^2 test is a t test that is used on more than 2 variables at a time. The student t test can also be used to compare 2 samples at a time, but if it is used to compare 5 samples, 2 at a time, the probability of obtaining a type one error is increased. That is, finding a significant difference when the two samples are the same. If a 5% error is used, the probability of obtaining such an error is $1 - 0.95^p$. Where p is the number of samples. Hotelling's T^2 is the preferred and recommended test method.

MANOVA (Multiple Analysis of Variance)

An analysis of variance is used for many independent X variables to solve one dependent Y variable. This method tests whether the mean differences among groups on a single dependent Y variable is significant.

For multiple independent X variables and multiple dependent Y factors, (that is, two or more Ys and one or more Xs), the multiple analysis of variance is used. MANOVA tests whether mean differences among groups of a combination of Ys are significant or not. The concept of various treatment levels and associated factors is still valid. The data should be normality distributed, have homogeneity of the covariance matrices, and have independence of observations.

In ANOVA, a sum of squares is used for the treatments and for the error term. In MANOVA the terms become matrices of the "sum of squares and cross-products." ANOVAs used multiple times across the dependent variables could result in inflated alpha errors. The MANOVA method is used to reduce the alpha risk by having only one test.

Refer to Table 8.15 for a further listing of differences between ANOVA and MANOVA.

MANOVA (Continued)

ANOVA	MANOVA	Description
Between sums of squares $SS_{between}$	SSCP (B)	Sum of the square and cross product matrix between groups
Within sums of squares SS_{within}	SSCP (W)	Sum of the square and cross product matrix within groups
Grand mean SS_{total}	SSCP (T)	Sum of the square and cross product matrix total
F test used for significance testing	Wilks' lambda λ	$\lambda = \frac{W}{T}$ (division of matrices) The variance not explained by the independent variable. $(1 - \lambda)^2 = \eta^2$ (eta-square) is equivalent to R^2 in regression.
	Lawley - Hotelling trace	These three other statistical tests provide similar results. Minitab displays all four tests.
	Roy's largest root	
	Pillai - Bartlett trace	

Table 8.15. ANOVA vs. MANOVA Differences

MANOVA Example

In an engineered plastics company, a multivariate experiment test was conducted having two independent variables (time and pressure of the extrusion process) at two levels, and three dependent responses (tensile strength, coefficient of friction, and bubble breaks). A MANOVA was conducted to test for relationships. The levels for the independent variables:

Time: high (+) equals 30 seconds, low (-) equals 10 seconds
Pressure: high (+) equals 80 psi, low (-) equals 20 psi

See Table 8.16 for the factors and levels.

Independent Variables		Responses		
Time	Pressure	Tensile strength	Coefficient of friction	Bubble breaks
-	-	10	48	39
+	-	31	46	62
-	+	12	22	45
+	+	32	28	62
-	-	14	44	32
+	-	33	42	69
-	+	16	22	44
+	+	34	24	63

Table 8.16. Experimental Independent Variables and Responses

The shortened Minitab output for the MANOVA is presented in Table 8.17. It only has the three statistics tables for the responses and interactions. Minitab automatically inserts the four statistical tests (Wilks', Lawley-Hotelling, Roy's, and Pillai-Bartlett) and makes the analysis. The results indicate that both the factors, time and pressure are significant with p values much below 5%. The interaction of time x pressure is not significant. For simplicity, the extensive SSCP tables were not displayed. For the individual familiar with linear algebra and matrices, the manual calculations can also be made.

MANOVA Example (Continued)

A MANOVA for tensile strength, coefficient of friction, and bubble breaks versus time and pressure is shown below:

MANOVA for Time s = 1 m = 0.5 n = 0.0				
Criterion	Test Statistic	F	DF	P
Wilks	0.00541	122.681	(3, 2)	0.008
Lawley-Hotelling	183.92659	122.681	(3, 2)	0.008
Pillai-Bartlett	0.99459	122.681	(3, 2)	0.008
Roy's	183.92659			

MANOVA for Pressure s = 1 m = 0.5 n = 0.0				
Criterion	Test Statistic	F	DF	P
Wilks	0.01291	50.988	(3, 2)	0.019
Lawley-Hotelling	76.48214	50.988	(3, 2)	0.019
Pillai-Bartlett	0.98709	50.988	(3, 2)	0.019
Roy's	76.48214			

MANOVA for Time x Pressure s = 1 m = 0.5 n = 0.0				
Criterion	Test Statistic	F	DF	P
Wilks	0.32009	1.416	(3, 2)	0.439
Lawley-Hotelling	2.12412	1.416	(3, 2)	0.439
Pillai-Bartlett	0.67991	1.416	(3, 2)	0.439
Roy's	2.1241			

Table 8.17. MANOVA Data Analysis

Hypothesis Testing

Hypothesis Testing is reviewed in the following topic areas:

- Terminology
- Significance
- Sample size
- Estimates
- Major Tests
- ANOVA
- Goodness-of-fit
- Contingency tables
- Nonparametric tests

Terminology

A number of commonly used hypothesis test terms are presented below.

Null Hypothesis

This is the hypothesis to be tested. The null hypothesis directly stems from the problem statement and is denoted as H_0. Examples:

- If one is investigating whether a modified seed will result in a different yield/acre, the null hypothesis (two-tail) would assume the yields to be the same H_o: $Y_a = Y_b$.

- If a strong claim is made that the average of process A is greater than the average of process B, the null hypothesis (one-tail) would state that process A $\leq$ process B. This is written as H_o: $A \leq B$.

The procedure employed in testing a hypothesis is strikingly similar to a court trial. The hypothesis is that the defendant is presumed not guilty until proven guilty. However, the term innocent does not apply to a null hypothesis. A null hypothesis can only be rejected, or fail to be rejected, it cannot be accepted because of a lack of evidence to reject it. If the means of two populations are different, the null hypothesis of equality can be rejected if enough data is collected. When rejecting the null hypothesis, the alternate hypothesis must be accepted.

Test Statistic

In order to test a null hypothesis, a test calculation must be made from sample information. This calculated value is called a test statistic and is compared to an appropriate critical value. A decision can then be made to reject or not reject the null hypothesis.

Terminology (Continued)

Types of Errors

When formulating a conclusion regarding a population based on observations from a small sample, two types of errors are possible:

- Type I error: This error occurs when the null hypothesis is rejected when it is, in fact, true. The probability of making a type I error is called α (alpha) and is commonly referred to as the producer's risk (in sampling). Examples are: incoming products are good but called bad; a process change is thought to be different when, in fact, there is no difference.

- Type II error: This error occurs when the null hypothesis is not rejected when it should be rejected. This error is called the consumer's risk (in sampling) and is denoted by the symbol β (beta). Examples are: incoming products are bad, but called good; an adverse process change has occurred but is thought to be no different.

The degree of risk (α) is normally chosen by the concerned parties (α is normally taken as 5%) in arriving at the critical value of the test statistic. The assumption is that a small value for α is desirable. Unfortunately, a small α risk increases the β risk. For a fixed sample size, α and β are inversely related. Increasing the sample size can reduce both the α and β risks.

The types of errors are shown in Figure 8.18 below:

		Null Hypothesis: True	Null Hypothesis: False
The Decision Made	Fail to Reject H_0	$p = 1 - \alpha$ Correct Decision	$p = \beta$ Type II Error
	Reject H_0	$p = \alpha$ Type I Error	$p = 1 - \beta$ Correct Decision

Figure 8.18 Error Matrix

Terminology (Continued)

One-Tail Test vs. Two-Tail Test

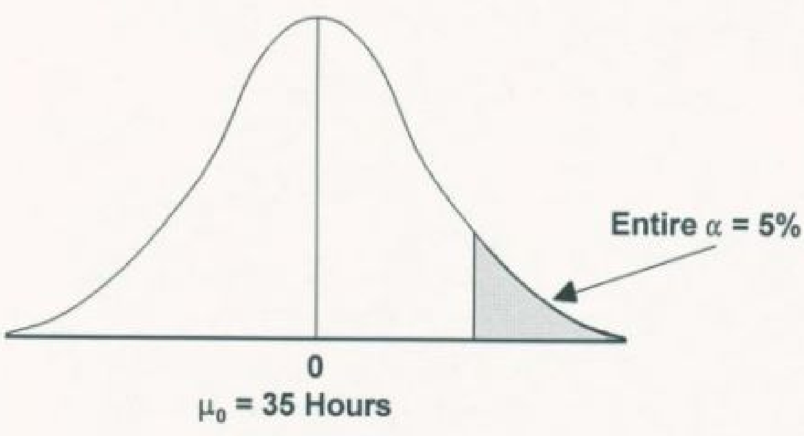
Any test of hypothesis has a risk associated with it and one is generally concerned with the α risk (a type I error which rejects the null hypothesis when it is true). The level of this α risk determines the level of confidence (1 - α) that one has in the conclusion. This risk factor is used to determine the critical value of the test statistic which is compared to a calculated value.

One-Tail Test

If a null hypothesis is established to test whether a sample value is smaller or larger than a population value, then the entire α risk is placed on one end of a distribution curve. This constitutes a one-tail test.

- A study was conducted to determine if the mean battery life produced by a new method is greater than the present battery life of 35 hours. In this case, the entire α risk will be placed on the right tail of the existing life distribution curve.

H_0: new < or = to present H_1: new > present

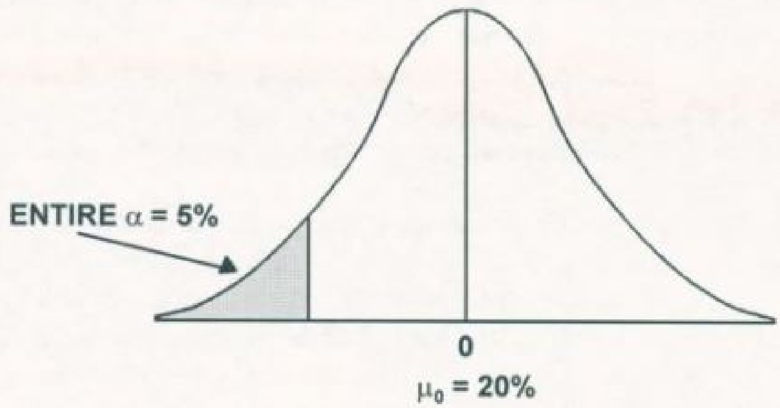


Determine if the true mean is within the α critical region.

- A chemist is studying the vitamin levels in a brand of cereal to determine if the process level has fallen below 20% of the minimum daily requirement. It is the manufacturer's intent to never average below the 20% level. A one-tail test would be applied in this case, with the entire α risk on the left tail.

Terminology (Continued)

H_0: level > or = 20% H_1: level < 20%



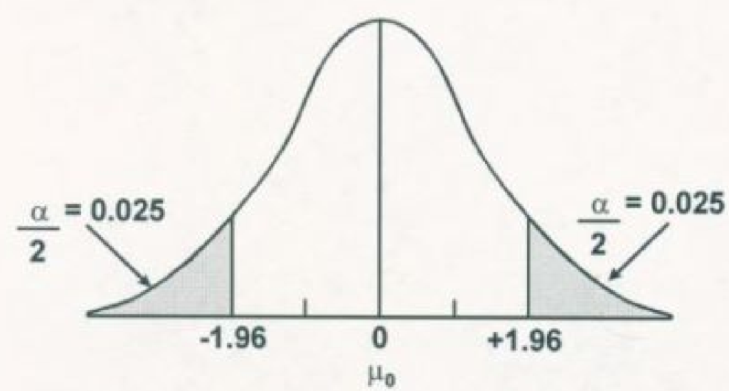
Determine if the true mean is within the α critical region.

Two-Tail Test

If a null hypothesis is established to test whether a population shift has occurred, in either direction, then a two-tail test is required. The allowable α error is generally divided into two equal parts. Examples:

- An economist must determine if unemployment levels have changed significantly over the past year.
- A study is made to determine if the salary levels of company A differ significantly from those of company B.

H_0: levels are = H_1: levels are ≠



Determine if the true mean is within either the upper or lower α critical regions.

Practical Significance vs. Statistical Significance

The hypothesis is tested to determine if a claim has significant statistical merit. Traditionally, levels of 5% or 1% are used for the critical significance values. If the calculated test statistic has a p-value below the critical level then it is deemed to be statistically significant. More stringent critical values may be required when human injury or catastrophic loss is involved. Less stringent critical values may be advantageous when there are no such risks and the potential economic gain is high.

On occasion, an issue of practical versus statistical significance may arise. That is, some hypothesis or claim is found to be statistically significant, but may not be worth the effort or expense to implement. This could occur if a large sample was tested to a certain value, such as a diet that results in a net loss of 0.5 pounds for 10,000 people. The result is statistically significant, but a diet losing 0.5 pounds per person would not have any practical significance. (Triola, 1994)[24]

Huck (1996)[9] indicates that issues of practical significance will often occur if the sample size is not adequate. A power analysis may be needed to aid in the decision-making process.

Power of Test H_0: $\mu = \mu_0$

Consider a null hypothesis that a population is believed to have mean $\mu_o = 70.0$ and $\sigma_{\bar{x}} = 0.80$. The 95% confidence limits are $70 \pm (1.96)(0.8) = 71.57$ and 68.43. One accepts the hypothesis $\mu = 70$ if $\bar{X}$s are between these limits. The alpha risk is that sample means will exceed those limits. One can ask "what if" questions such as, "What if μ shifts to 71, would it be detected?" There is a risk that the null hypothesis would be accepted even if the shift occurred. This risk is termed β.

Power of Test H_0: $\mu = \mu_0$ (Continued)

The value of β is large if μ is close to μ_0 and small if μ is very different from μ_0. This indicates that slight differences from the hypothesis will be difficult to detect and large differences will be easier to detect. The normal distribution curves below show the null and alternative hypotheses. If the process shifts from 70 to 71, there is a 76% probability that it would not be detected.

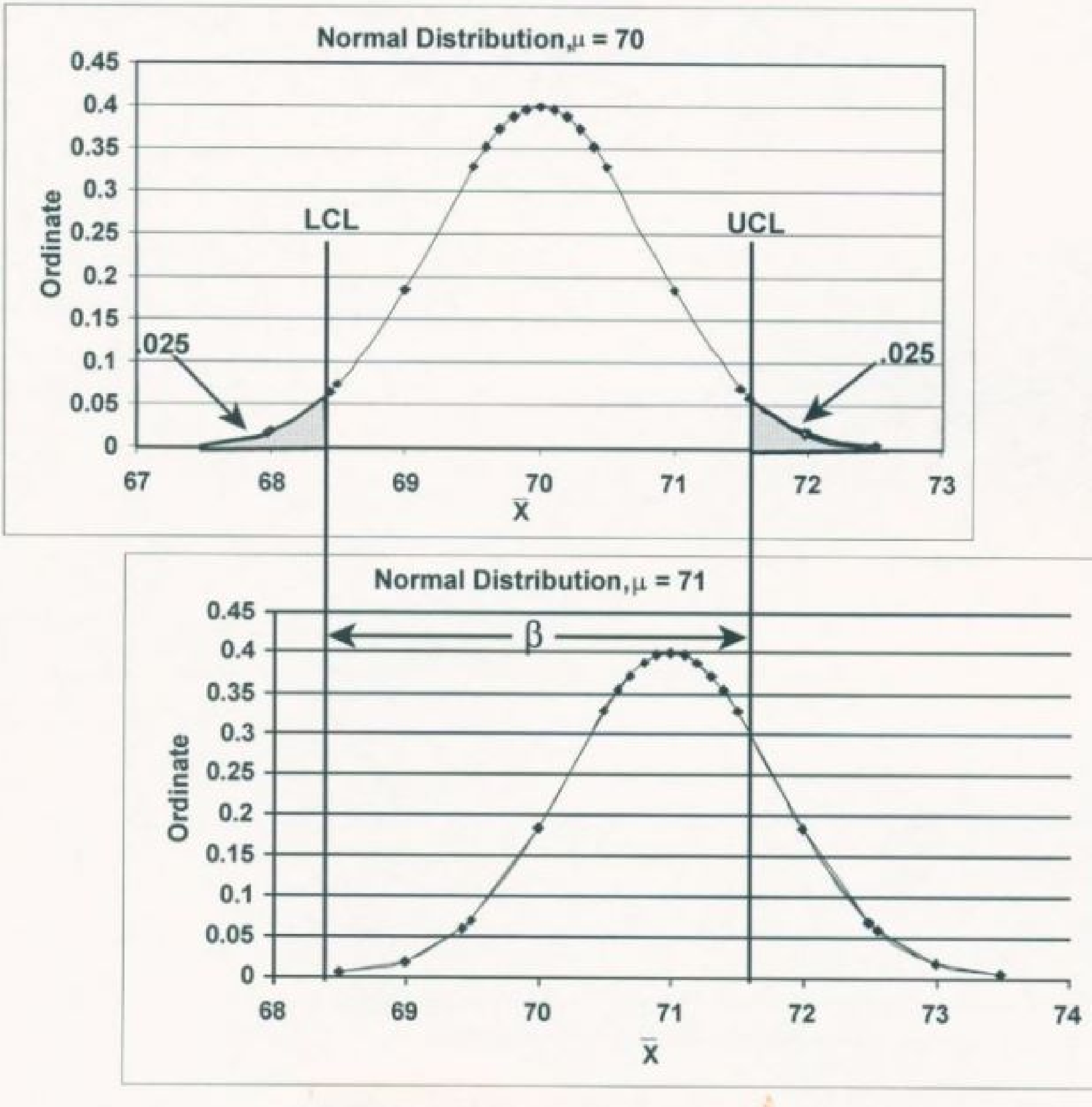


Figure 8.19 Illustration of Beta (β) Risk

Power of Test H_0: $\mu = \mu_0$ (Continued)

To construct a power curve, $1 - \beta$ is plotted against alternative values of μ. The power curve for the process under discussion is shown below. A shift in a mean away from the null increases the probability of detection. In general, as alpha increases, beta decreases and the power of $1 - \beta$ increases.

One can say that a gain in power can be obtained by accepting a lower level of protection from the alpha error. Increasing the sample size makes it possible to decrease both alpha and beta and increase power.

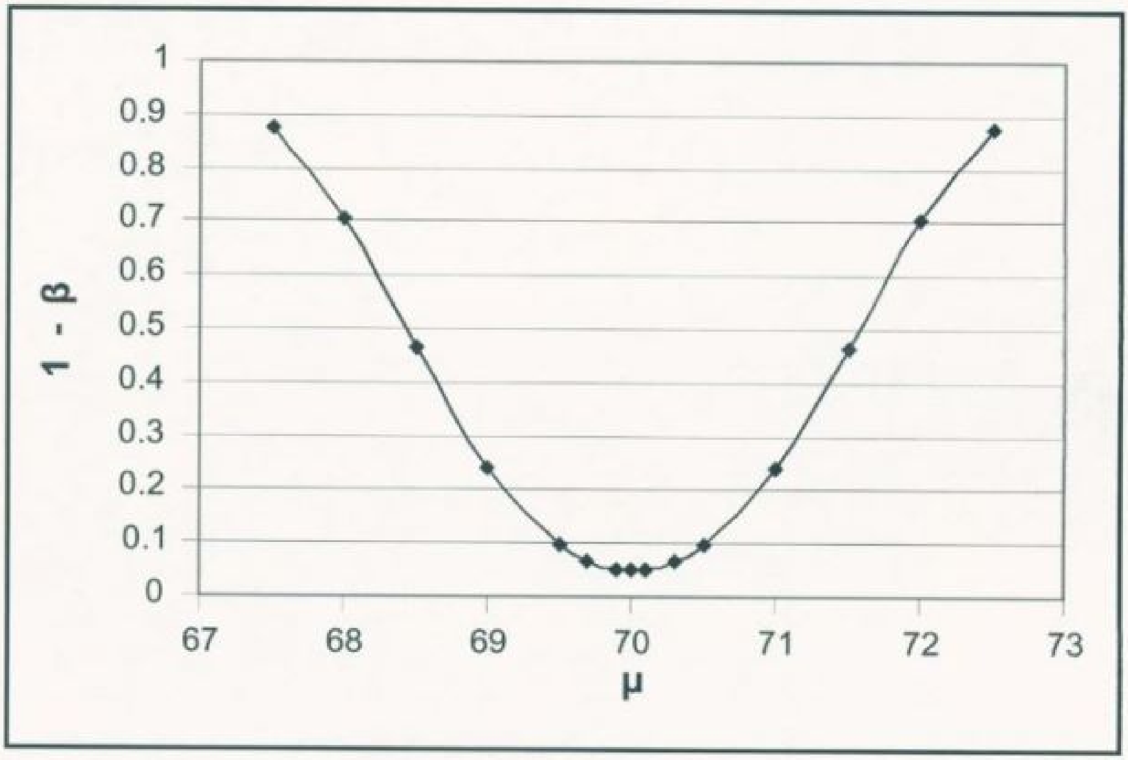


Figure 8.20 Power Curve, $(1 - \beta)$ vs μ

The concept of power also relates to experimental design and analysis of variance. The following equation briefly states the relationship for ANOVA.

$$1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ is false})$$

$1 - \beta$ = Probability of rejecting the null hypothesis given that the null hypothesis is false.

Sample Size

In the statistical inference discussion thus far, it has been assumed that the sample size (n) for hypothesis testing has been given and that the critical value of the test statistic will be determined based on the α error that can be tolerated. The ideal procedure, however, is to determine the α and β error desired and then to calculate the sample size necessary to obtain the desired decision confidence.

The sample size (n) needed for hypothesis testing depends on:

- The desired type I (α) and type II (β) risk
- The minimum value to be detected between the population means ($\mu - \mu_0$)
- The variation in the characteristic being measured (S or σ)

Variable data sample size, only using α, is illustrated by the following:

Example 8.7: Assume in a pilot process one wishes to determine whether an operational adjustment will alter the process hourly mean yield by as much as 4 tons per hour. What is the minimum sample size which, at the 95% confidence level (Z=1.96), would confirm the significance of a mean shift greater than 4 tons per hour? Historic information suggests that the standard deviation of the hourly output is 20 tons. The general sample size equation for variable data (normal distribution) is:

$$n = \frac{Z^2\,\sigma^2}{E^2} = \frac{(1.96)^2(20)^2}{(4)^2} = 96$$

Obtain 96 pilot hourly yield values and determine the hourly average. If this mean deviates by more than 4 tons from the previous hourly average, a significant change at the 95% confidence level has occurred. If the sample mean deviates by less than 4 tons/hr, the observable mean shift can be explained by chance cause.

For binomial data, use the following formula:

$$n = \frac{Z^2(\bar{p})(1 - \bar{p})}{(\Delta p)^2}$$

Where,

Z = The appropriate Z value
Δp = The desired proportion interval
$\bar{p}$ = Proportion rate
n = Sample size

Estimators

In analyzing sample values to arrive at population probabilities, two major estimators are used: point estimation and interval estimation.

Example 8.8: Consider the following tensile strength readings from 4 piano wire segments: 28.7, 27.9, 29.2 and 26.5 psi. Based on this data, the following expressions are true:

- Point estimation: If a single estimate value is desired (i.e., the sample average), then a point estimate represented by $\bar{X}$ can be obtained.

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{28.7+27.9+29.2+26.5}{4} = 28.08 \text{ and } (s = 1.1786)$$

28.08 psi is the point estimate for the population mean.

- Interval Estimate or CI (Confidence Interval): From sample data one can calculate the interval within which the population mean is predicted to fall. Confidence intervals are always estimated for population parameters and, in general, are derived from the mean and standard deviation of sample data. For small samples, a critical value from the t distribution is required and for 95% confidence, t = 3.182 for n-1 degrees of freedom. The CI equation and interval would be:

$$\bar{X} \pm 3.182\frac{s}{\sqrt{n}} = 28.08 \pm 3.182\left(\frac{1.1786}{2}\right) = 26.205 \text{ and } 29.955$$

If the population sigma is known (say σ = 2 psi), the Z distribution is used. The critical Z value for 95% confidence is 1.96. The CI equation and interval would be:

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = 28.08 \pm 1.96\left(\frac{2}{2}\right) = 26.12 \text{ and } 30.04$$

A confidence interval is a two-tail event and requires critical values based on an alpha/2 risk in each tail. The central limit theorem term, $\sigma/\sqrt{n}$, is necessary because the confidence interval is for a population mean and not individual values.

Note that other confidence interval formulas exist. These include percent nonconforming, Poisson distribution data and very small sample size data.

Confidence Intervals for the Mean

Continuous Data - Large Samples

Use the normal distribution to calculate the confidence interval for the mean.

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Where:
$\bar{X}$ = the sample average
σ = the population standard deviation
n = the sample size
$Z_{\frac{\alpha}{2}}$ = the normal distribution value for a desired confidence level

Example 8.9: The average of 100 samples is 18 with a population standard deviation of 6. Calculate the 95% confidence interval for the population mean.

$$\mu = 18 \pm 1.96 \frac{(6)}{\sqrt{100}} = 18 \pm 1.176$$

$$16.82 \le \mu \le 19.18$$

Continuous Data - Small Samples

If a relatively small sample is used (<30) then the t distribution must be used.

$$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Where:
$\bar{X}$ = the sample average
s = the sample standard deviation
n = the sample size
$t_{\frac{\alpha}{2}}$ = the t distribution value for a desired confidence level and (n - 1) degrees of freedom

Example 8.10: Use the same values as in the prior example except that the sample size is 25.

$$\mu = 18 \pm 2.064 \frac{(6)}{\sqrt{25}} = 18 \pm 2.48$$

$$15.52 \le \mu \le 20.48$$

Confidence Intervals for Variation

The confidence intervals for the mean were symmetrical about the average. This is not true for the variance, since it is based on the chi square distribution. The formula is:

$$\frac{(n-1)s^2}{X^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{X^2_{1-\frac{\alpha}{2},n-1}}$$

Where: n = the sample size
s^2 = point estimate of variance
$X^2_{\frac{\alpha}{2}}$ and $X^2_{1-\frac{\alpha}{2}}$ = are the table values for (n - 1) degrees of freedom

Example 8.11: The sample variance for a set of 25 samples was found to be 36. Calculate the 90% confidence interval for the population variance.

$$\frac{(24)(36)}{36.42} \le \sigma^2 \le \frac{(24)(36)}{13.85}$$

$$23.72 \le \sigma^2 \le 62.38$$

Confidence Intervals for Proportion

For large sample sizes, with n(p) and n(1-p) greater than or equal to 4 or 5, the normal distribution can be used to calculate a confidence interval for proportion. The following formula is used:

$$p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

Where: p = the population proportion estimate
n = the sample size
$Z_{\frac{\alpha}{2}}$ = the appropriate confidence level from a Z table

Example 8.12: If 16 defectives were found in a s
ample size of 200 units, calculate the 90% confidence interval for the proportion.

$$0.08 \pm 1.645 \sqrt{\frac{(0.08)(0.92)}{200}} = 0.08 \pm 0.032$$

$$0.048 \le p \le 0.112$$

Hypotheses Tests

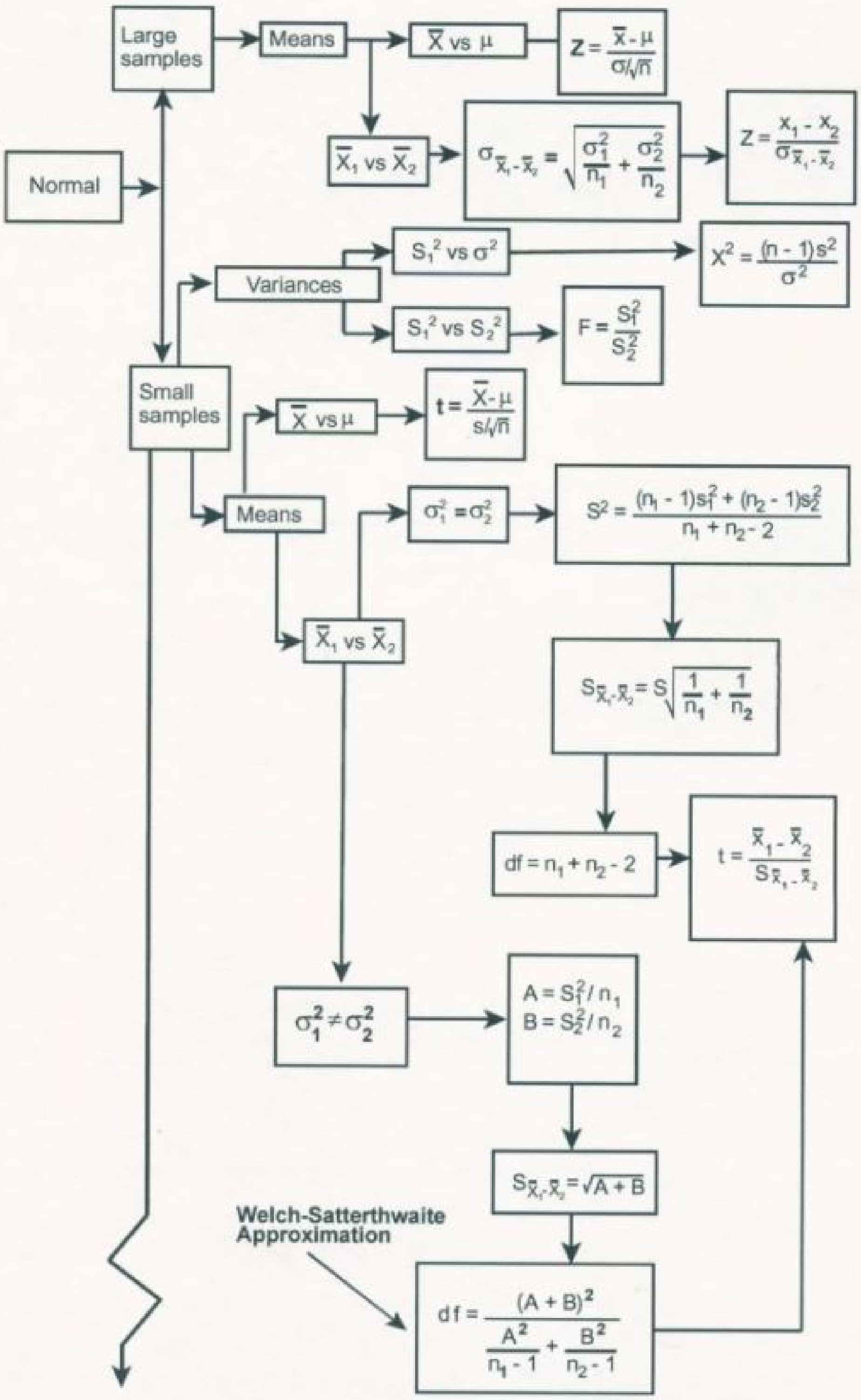
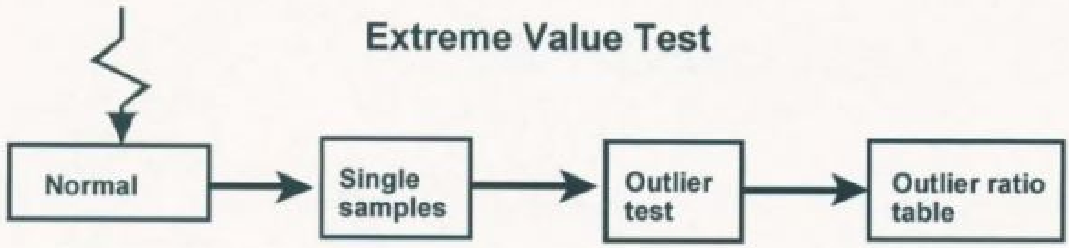


Figure 8.21 Schematic of Normal Distribution Hypotheses

Hypotheses Tests (Continued)

Extreme Value Test



Attribute Distribution Hypotheses

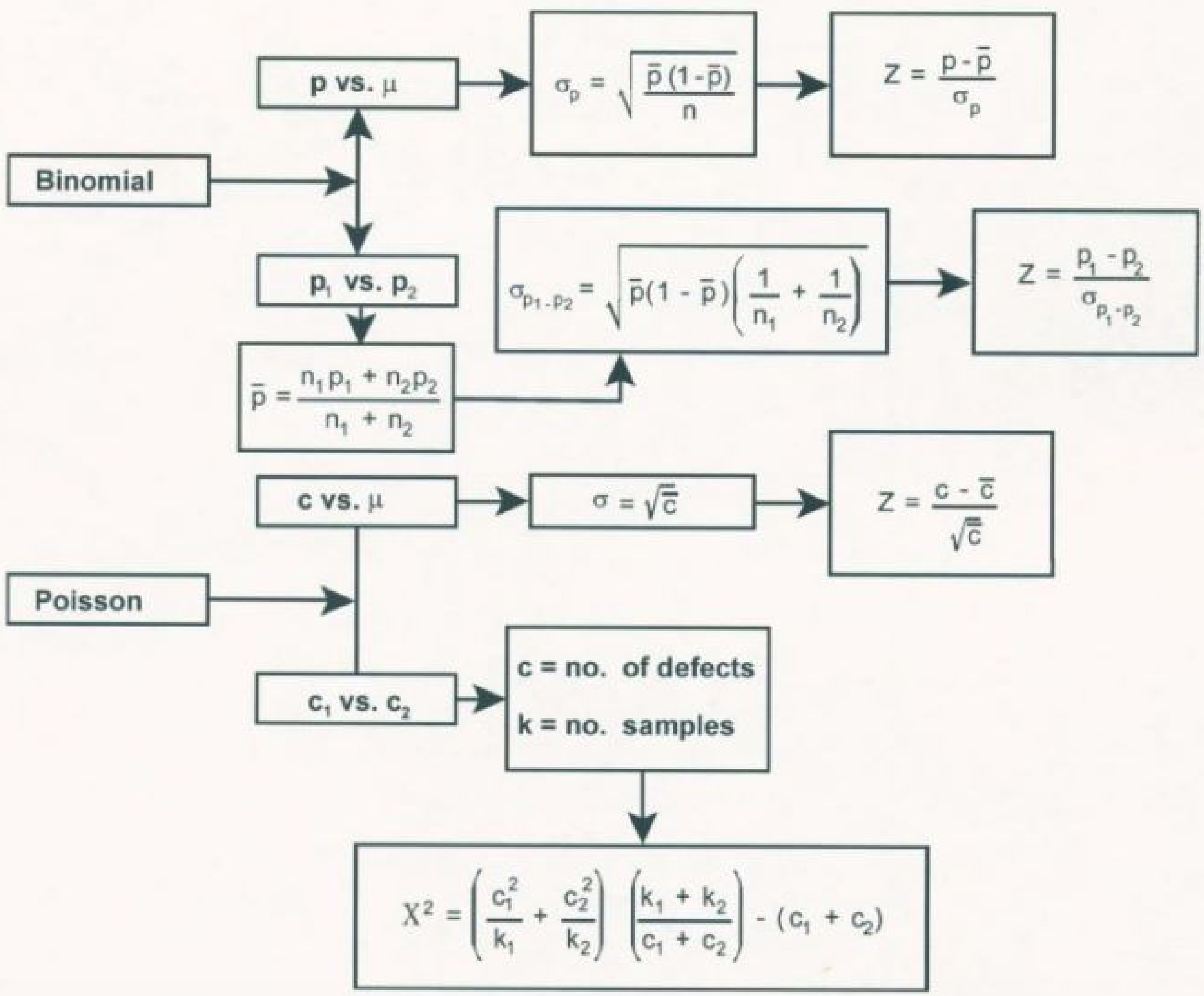


Figure 8.22 Schematic of Attribute Distribution Hypotheses

The above charts are provided by DuWayne Carlson.

Hypotheses Tests for Means

Z Test

When the population follows a normal distribution and the population standard deviation, σ_X, is known, then the hypothesis tests for comparing a population mean, μ, with a fixed value, μ_0, are given by the following:

$H_0: \mu = \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu \geq \mu_0$
$H_1: \mu \neq \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$

The null hypothesis is denoted by H_0 and the alternative hypothesis is denoted by H_1. The test statistic is given by:

$$\boxed{Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\sigma_X / \sqrt{n}}}$$

... where the sample average is $\bar{X}$, the number of samples is n and the standard deviation of the mean is $\sigma_{\bar{X}}$. Note, if $n > 30$, that the sample standard deviation, s, is often used as an estimate of the population standard deviation, σ_X. The test statistic, Z, is compared with a critical value Z_α or $Z_{\alpha/2}$ which is based on a significance level, α, for a one-tailed test or $\alpha/2$ for a two-tailed test. If the H_1 sign is $\neq$, it is a two-tailed test. If the H_1 sign is >, it is a right, one-tailed test, and if the H_1 sign is <, it is a left, one-tailed test.

(Triola, 1994)[24]

Example 8.13: The average vial height from an injection molding process has been 5.00" with a standard deviation of 0.12". An experiment is conducted using new material which yielded the following vial heights: 5.10", 4.90", 4.92", 4.87", 5.09", 4.89", 4.95", and 4.88".

Can one state with 95% confidence that the new material is producing shorter vials with the existing molding machine setup? This question involves an inference about a population mean with a known sigma. The Z test applies. The null and alternative hypotheses are:

$H_0: \mu \geq \mu_0$ $\quad$ $H_1: \mu < \mu_0$

or

$H_0: \mu \geq 5.00"$ $\quad$ $H_1: \mu < 5.00"$

Hypotheses Tests for Means (Continued)

Z Test (Continued)

Example 8.13 (continued):

The sample average is $\bar{X} = 4.95"$ with n = 8 and the population standard deviation is $\sigma_X = 0.12"$. The test statistic is:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_X / \sqrt{n}} = \frac{4.95 - 5.00}{0.12 / \sqrt{8}} = \frac{-0.05}{0.042} = -1.18$$

Since the H_1 sign is <, it is a left, one-tailed test and with a 95% confidence, the level of significance, $\alpha = 1 - 0.95 = 0.05$. Looking up the critical value in a normal distribution or Z table, one finds $Z_{0.05} = -1.645$. Since the test statistic, -1.18, does not fall in the reject (or critical) region, the null hypothesis cannot be rejected. There is insufficient evidence to conclude that the vials made with the new material are shorter.

If the test statistic had been, for example -1.85, we would have rejected the null hypothesis and concluded the vials made with the new material are shorter.

Student's t Test

The student's t distribution applies to samples drawn from a normally distributed population. It is used for making inferences about a population mean when the population variance, σ^2, is unknown and the sample size, n, is small. The use of the t distribution is never wrong for any sample size. However, a sample size of 30 is normally the crossover point between the t and Z tests. The test statistic formula is:

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

Where: $\bar{X}$ = Sample mean
μ_0 = Target value or population mean
s = Sample standard deviation
n = Number of test samples

The null and alternative hypotheses are the same as were given for the Z test.

Hypotheses Tests for Means (Continued)

Student's t Test (Continued)

The test statistic, t, is compared with a critical value, t_α or $t_{\alpha/2}$, which is based on a significance level, α, for a one-tailed test or α/2 for a two-tailed test, and the number of degrees of freedom, d.f. The degrees of freedom is determined by the number of samples, n, and is simply:

$$\text{d.f.} = n - 1$$

Example 8.14: The average daily yield of a chemical process has been 880 tons (μ = 880 tons). A new process has been evaluated for 25 days (n = 25) with a yield of 900 tons ($\bar{X}$) and sample standard deviation, s = 20 tons. Can one say with 95% confidence that the process has changed?

The null and alternative hypotheses are:

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

or

$$H_0: \mu = 880 \text{ tons} \qquad H_1: \mu \neq 880 \text{ tons}$$

The test statistic calculation is:

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{900 - 880}{20 / \sqrt{25}} = \frac{20}{4} = 5$$

Since the H_1 sign is ≠, it is a two-tailed test and with a 95% confidence, the level of significance, α = 1 - 0.95 = 0.05. Since it is a two-tail test, α/2 is used to determine the critical values. The degrees of freedom, d.f. = n - 1 = 24. Looking up the critical values in a t distribution table, one finds $t_{0.025} = -2.064$ and $t_{0.975} = 2.064$. Since the test statistic, 5, falls in the right-hand reject (or critical) region, the null hypothesis is rejected. We conclude with 95% confidence that the process has changed.

This technique was developed by W. S. Gosset and published in 1908 under the pen name "Student." Gosset referred to the quantity under study as t. The test has since been known as the student's t test.

Hypotheses Tests for Means (Continued)

Student's t Test (Continued)

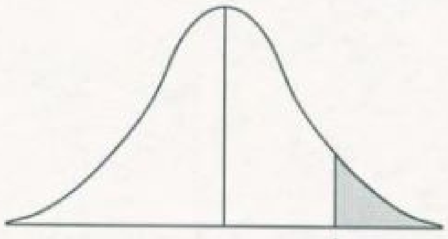
Example 8.15: A new spark plug design is tested for wear. A sample of six plugs yielded: 0.0058", 0.0049", 0.0052", 0.0044", 0.0050" and 0.0047" of wear. The current design has historically produced an average wear of 0.0055". With 95% confidence is the new design better?

Example 8.16: A very expensive experiment has been conducted to evaluate the manufacture of synthetic diamonds by a new technique. Five diamonds have been generated with recorded weights of 0.46, 0.61, 0.52, 0.57 and 0.54 carats. An average diamond weight equal to or greater than 0.50 carats must be realized for the venture to be profitable. What is your recommendation assuming 95% confidence?

STEPS	EXAMPLE 8.15	EXAMPLE 8.16
Step 1: Establish the null hypothesis: [there is no difference between the target value and the sample average]	H_0: $\mu_1 \geq 0.0055"$ H_1: $\mu_1 < 0.0055"$	H_0: $\mu_2 \leq 0.50$ H_1: $\mu_2 > 0.50$
Step 2: Determine the critical value of t for a 95% confidence level from the t distribution	DF = 5 from (n - 1)	DF = 4 from (n - 1)
	Left Tail	Right Tail
	-2.015	2.132
Step 3: Calculate the t statistic: $t = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$	$t = \dfrac{0.0050 - 0.0055}{0.00048/\sqrt{6}}$ t = -2.551	$t = \dfrac{0.54 - 0.50}{0.056/\sqrt{5}}$ t = 1.597
Step 4: Can we reject the null hypothesis?	Since the value of calculated t is to the left of -2.015, the null hypothesis is rejected. The wear is less for the new plug design.	Since the value of calculated t (1.597) is not to the right of the critical t (2.132), the null hypothesis can't be rejected. Insufficient evidence exists for the new technique to be profitable.

Table 8.23 A Matrix Review of Two Student's t Tests

One underlying assumption is that the sampled population has a normal probability distribution. This is a restrictive assumption since the distribution of the sample is unknown. The t distribution works well for distributions that are bell-shaped.



t Distribution Table

d.f.	$t_{0.100}$	$t_{0.050}$*	$t_{0.025}$**	$t_{0.010}$	$t_{0.005}$	d.f.
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

* One-tail 5% α risk ** Two-tail 5% α risk

Table 8.24 Student's t Distribution Table

There is only a 5% probability that a sample with 10 degrees of freedom will have a t value greater than 1.812.

Hypothesis Tests for Variance

Chi Square (X^2) Test

Standard deviation (or variance) is fundamental in making inferences regarding the population mean. In many practical situations, variance (σ^2) assumes a position of greater importance than the population mean. Consider the following examples:

1. A shoe manufacturer wishes to develop a new sole material with a more stable wear pattern. The wear variation in the new material must be smaller than the variation in the existing material.

2. An aircraft altimeter manufacturer wishes to compare the measurement precision among several instruments.

3. Several inspectors examine finished parts at the end of a manufacturing process. Even when the same lots are examined by different inspectors, the number of defectives varies. Their supervisor wants to know if there is a significant difference in the knowledge or abilities of the inspectors.

The above problems represent a comparison of a target or population variance with an observed sample variance, a comparison between several sample variances, or a comparison between frequency proportions. The standardized test statistic is called the chi square (X^2) test.

Population variances are distributed according to the chi square distribution. Therefore, inferences about a single population variance will be based on chi square.

The chi square test is widely used in two applications.

Case I. Comparing variances when the variance of the population is known.

Case II. Comparing observed and expected frequencies of test outcomes when there is no defined population variance (attribute data).

Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

When the population follows a normal distribution, the hypothesis tests for comparing a population variance, σ_X^2, with a fixed value, σ_0^2, are given by the following:

H_0: $\sigma_X^2 = \sigma_0^2$	H_0: $\sigma_X^2 \le \sigma_0^2$	H_0: $\sigma_X^2 \ge \sigma_0^2$
H_1: $\sigma_X^2 \neq \sigma_0^2$	H_1: $\sigma_X^2 > \sigma_0^2$	H_1: $\sigma_X^2 < \sigma_0^2$

The null hypothesis is denoted by H_0 and the alternative hypothesis is denoted by H_1. The test statistic is given by:

$$X^2 = \frac{(n-1)s^2}{\sigma_x^2}$$

Where the number of samples is n and the sample variance is s^2. The test statistic, X^2, is compared with a critical value X^2_α or $X^2_{\alpha/2}$ which is based on a significance level, α, for a one-tailed test or $\alpha/2$ for a two-tailed test and the number of degrees of freedom, d.f. The degrees of freedom is determined by the number of samples, n, and is simply:

$$\text{d.f.} = n - 1$$

If the H_1 sign is $\neq$, it is a two-tailed test. If the H_1 sign is >, it is a right, one-tailed test, and if the H_1 sign is <, it is a left, one-tailed test. (Triola, 1994)[24]

The X^2 distribution looks like so:

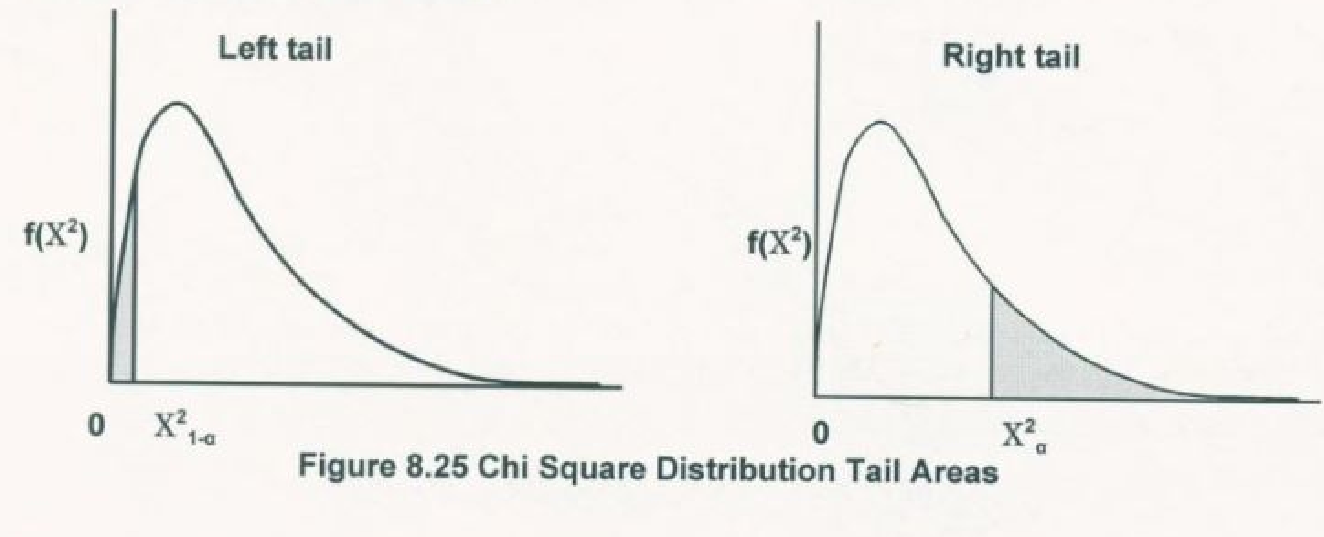


Figure 8.25 Chi Square Distribution Tail Areas

Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

The critical values of the chi square distribution are shown in a simplified Table on the following page. Please note, unlike the Z and t distributions, the tails of the chi square distribution are non-symmetrical.

Chi square Case I. Comparing Variances When the Variance of the Population Is Known.

Example 8.17: The R & D department of a steel plant has tried to develop a new steel alloy with less tensile variability. The R & D department claims that the new material will show a four sigma tensile variation less than or equal to 60 psi 95% of the time. An eight sample test yielded a standard deviation of 8 psi. Can a reduction in tensile strength variation be validated with 95% confidence?

Solution: The best range of variation expected is 60 psi. This translates to a sigma of 15 psi (an approximate 4 sigma spread covering 95.44% of occurrences).

The null hypothesis is: $H_0: \sigma_1^2 \geq (15)^2$

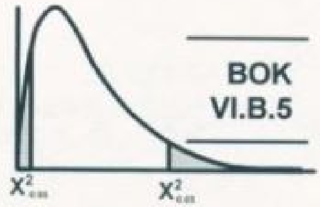
The alternative hypothesis is: $H_1: \sigma_1^2 < (15)^2$

From the chi square table: Because S is less than σ, this is a left tail test with $n - 1 = 7$. The critical value for 95% confidence is 2.17. That is, the calculated value will be less than 2.17, 5% of the time. Please note that if one were looking for more variability in the process a right tail rejection region would have been selected and the critical value would be 14.07.

The calculated statistic is:

$$X^2 = \frac{(n-1)s^2}{\sigma_1^2} = \frac{(7)(8)^2}{(15)^2} = 1.99$$

Since 1.99 is less than 2.17, the null hypothesis must be rejected. The decreased variation in the new steel alloy tensile strength supports the R & D claim.



Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

DF	$X^2_{0.99}$	$X^2_{0.95}$	$X^2_{0.90}$	$X^2_{0.10}$	$X^2_{0.05}$	$X^2_{0.01}$
1	0.00016	0.0039	0.0158	2.71	3.84	6.63
2	0.0201	0.1026	0.2107	4.61	5.99	9.21
3	0.115	0.352	0.584	6.25	7.81	11.34
4	0.297	0.711	1.064	7.78	9.49	13.28
5	0.554	1.15	1.61	9.24	11.07	15.09
6	0.872	1.64	2.20	10.64	12.59	16.81
7	1.24	2.17	2.83	12.02	14.07	18.48
8	1.65	2.73	3.49	13.36	15.51	20.09
9	2.09	3.33	4.17	14.68	16.92	21.67
10	2.56	3.94	4.87	15.99	18.31	23.21
11	3.05	4.57	5.58	17.28	19.68	24.73
12	3.57	5.23	6.30	18.55	21.03	26.22
13	4.11	5.89	7.04	19.81	22.36	27.69
14	4.66	6.57	7.79	21.06	23.68	29.14
15	5.23	7.26	8.55	22.31	25.00	30.58
16	5.81	7.96	9.31	23.54	26.30	32.00
18	7.01	9.39	10.86	25.99	28.87	34.81
20	8.26	10.85	12.44	28.41	31.41	37.57
24	10.86	13.85	15.66	33.20	36.42	42.98
30	14.95	18.49	20.60	40.26	43.77	50.89
40	22.16	26.51	29.05	51.81	55.76	63.69
60	37.48	43.19	46.46	74.40	79.08	88.38
120	86.92	95.70	100.62	140.23	146.57	158.95

Table 8.26 Chi Square Critical Values

The above table addresses both the left and right tails of X^2. 95% of the area under the X^2 distribution lies to the right of $X^2_{0.95}$. Example: For a right tail evaluation with d.f. = 5, only 5% of the values will exceed the critical value of 11.07.

Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

Chi square Case II. Comparing Observed and Expected Frequencies of Test Outcomes. (Attribute Data)

It is often necessary to compare proportions representing various process conditions. Machines may be compared as to their ability to produce precise parts. The ability of inspectors to identify defective products can be evaluated. This application of chi square is called the contingency table or row and column analysis.

The procedure is as follows:

1. Take one subgroup from each of the various processes and determine the observed frequencies (O) for the various conditions being compared.

2. Calculate for each condition the expected frequencies (E) under the assumption that no differences exist among the processes.

3. Compare the observed and expected frequencies to obtain "reality." The following calculation is made for each condition:

$$\frac{(O - E)^2}{E}$$

4. Total all the process conditions:

$$X^2 = \sum \frac{(O - E)^2}{E}$$

This is the most "famous" chi square statistic.

5. A critical value is determined using the chi square table with the entire level of significance, α, in the one-tail, right side, of the distribution. The degrees of freedom is determined from the calculation (R-1)(C-1) [the number of rows minus 1 times the number of columns minus 1].

6. A comparison between the test statistic and the critical value confirms if a significant difference exists (at a selected confidence level).

Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

Example 8.18: An airport authority wanted to evaluate the ability of three X-ray inspectors to detect key items. A test was devised whereby transistor radios were placed in ninety pieces of luggage. Each inspector was exposed to exactly thirty of the preselected and "bugged" items in a random fashion. The observed results are summarized below.

	Inspectors			Treatment Totals
	1	2	3	
Radios detected	27	25	22	74
Radios undetected	3	5	8	16
Sample total	30	30	30	90

Is there any significant difference in the abilities of the inspectors? (95% confidence)

Null hypothesis:

There is no difference among the three inspectors, H_0: $p_1 = p_2 = p_3$

Alternative hypothesis:

At least one of the proportions is different, H_1: $p_1 \neq p_2 \neq p_3$

The degrees of freedom = (rows - 1)(columns - 1) = (2-1)(3-1) = 2

The critical value of X^2 for DF = 2 and α = 0.05 in the one-tail, right side of the distribution, is 5.99 (refer to Table 8.26 or the Appendix Section XII). There is only a 5% chance that the calculated value of X^2 will exceed 5.99.

Hypothesis Tests for Variance (Continued)

Chi Square (X^2) Test (Continued)

Example 8.18 (continued):
The expected values for the test are shown below:

	1	2	3	Treatment Totals
Radios detected	24.67	24.67	24.67	74
Radios undetected	5.33	5.33	5.33	16
Sample total	30	30	30	90

$$\text{The expected values} = \frac{\text{row total x column total}}{\text{grand total}}$$

$$X^2 = \sum \frac{(O - E)^2}{E}$$

$$X^2 = \frac{(2.33)^2}{24.67} + \frac{(0.33)^2}{24.67} + \frac{(2.67)^2}{24.67} + \frac{(2.33)^2}{5.33} + \frac{(0.33)^2}{5.33} + \frac{(2.67)^2}{5.33}$$

$$X^2 = 0.220 + 0.004 + 0.289 + 1.019 + 0.020 + 1.338$$

$$X^2 = 2.89$$

Since the calculated value of X^2 is less than the previously calculated critical value of 5.99 and this is a right tail test, the null hypothesis cannot be rejected. There is insufficient evidence to say with 95% confidence that the abilities of the inspectors differ.

Hypothesis Tests for Proportions

p Test

When testing a claim about a population proportion, with a fixed number of independent trials having constant probabilities, and each trial has two outcome possibilities (a binomial experiment), a p test can be used. When $np < 5$ or $n(1-p) < 5$, the binomial distribution is used to test hypotheses relating to proportion.

If conditions that $np \geq 5$ and $n(1-p) \geq 5$ are met, then the binomial distribution of sample proportions can be approximated by a normal distribution. The hypothesis tests for comparing a sample proportion, p, with a fixed value, p_0, are given by the following:

$H_0: p = p_0$	$H_0: p \leq p_0$	$H_0: p \geq p_0$
$H_1: p \neq p_0$	$H_1: p > p_0$	$H_1: p < p_0$

The null hypothesis is denoted by H_0 and the alternative hypothesis is denoted by H_1. The test statistic is given by:

$$Z = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}}$$

Where the number of successes is x and the number of samples is n. The test statistic, Z, is compared with a critical value Z_α or $Z_{\alpha/2}$ which is based on a significance level, α, for a one-tailed test or α/2 for a two-tailed test. If the H_1 sign is ≠, it is a two-tailed test. If the H_1 sign is >, it is a right, one-tailed test, and if the H_1 sign is <, it is a left, one-tailed test.

(Triola, 1994)[24]

Paired-Comparison Hypotheses Tests

2 Mean, Equal Variance t Test

The 2 mean, equal variance t test examines the difference between 2 sample means ($\bar{X}_1$ vs $\bar{X}_2$) when σ_1 and σ_2 are unknown but considered equal.

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

s_p = Pooled standard deviation

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

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad DF = n_1 + n_2 - 2$$

Example 8.19: Compare product weight data from two machines.

Machine 1	Machine 2
3.125	3.110
3.120	3.095
3.135	3.115
3.130	3.120
3.125	3.125

$\bar{X}_1 = 3.127$ $\quad$ $\bar{X}_2 = 3.113$

$s_1 = 0.0057$ $\quad$ $s_2 = 0.0115$

$$s_p = \sqrt{\frac{4(0.0057)^2 + 4(0.0115)^2}{8}}$$

$$t = \frac{3.127 - 3.113}{0.0091\sqrt{\frac{1}{5} + \frac{1}{5}}} = 2.43 \qquad DF = 5 + 5 - 2 = 8$$

The critical value for $t_{0.025,\,8} = 2.306$ (a 2 sided test for $\alpha = 0.05$)
The null hypothesis H_0 is rejected.

Paired-Comparison Hypotheses Tests (Continued)

2 Mean, Unequal Variance t Test

The 2 mean, unequal variance t test examines the difference between 2 sample means ($\bar{X}_1$ vs $\bar{X}_2$) when σ_1 and σ_2 are unknown, but are not considered equal.

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$DF = \frac{1}{\frac{\left(\frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)^2}{n_1 - 1} + \frac{\left(\frac{\frac{s_2^2}{n_2}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)^2}{n_2 - 1}}$$

Example 8.20: Use data from the prior example to perform an unequal variance t test.

$$DF = 5.83 = 5$$

Note, if DF is rounded off conservatively, the effect is to increase the confidence level rather than to reduce it.

$$t = \frac{3.127 - 3.113}{\sqrt{\frac{(0.0057)^2}{5} + \frac{(0.0115)^2}{5}}} = 2.440$$

The critical value for $t_{0.025,5} = 2.571$ (two sided test for $\alpha = 0.05$). The null hypothesis H_0 cannot be rejected (although it is fairly close).

Paired-Comparison Hypotheses Tests (Continued)

Paired t Test

The paired t test examines the difference between 2 sample means. Data is taken in pairs with the difference calculated for each pair.

Example 8.21: Suppose, in the previous problem, we had 2 operators measure the same set of samples. Calculate a paired t test.

$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$

Sample	Operator 1	Operator 2	Difference (d)
1	3.125	3.11	0.015
2	3.12	3.095	0.025
3	3.135	3.115	0.02
4	3.13	3.12	0.01
5	3.125	3.125	0
			$\bar{d} = 0.014$ *
			$s_d = 0.0096$

DF = n - 1 = 4

* The sign of $\bar{d}$ must be maintained

$$t = \frac{\bar{d}}{\left(\frac{s_d}{\sqrt{n}}\right)} = \frac{0.014}{\left(\frac{0.0096}{\sqrt{5}}\right)} = 3.261$$

$t_{0.025,\,4} = 2.776$ (two sided test for $\alpha = 0.05$; a paired t test is usually a two-tail test)

The null hypothesis H_0 is rejected.

The paired method (dependent samples), compared to treating the data as two independent samples, will often show a more significant difference because the standard deviation of the differences (S_d) includes no sample-to-sample variation. This comparatively more frequent significance occurs despite the fact that "n - 1" represents fewer degrees of freedom than "$n_1 + n_2$ -2." In general, the paired t test is a more sensitive test than the comparison of two independent samples.

Paired-Comparison Hypotheses Tests (Continued)

F Test

The need for a statistical method of comparing two population variances is apparent. One may wish to compare the precision of one measuring device to another, or the relative stability of two manufacturing processes. The F test named in honor of Sir Ronald Fisher is usually employed. APP.

If independent random samples are drawn from two normal populations with equal variances, the ratio of $(S_1^2)/(S_2^2)$ creates a sampling distribution known as the F distribution. The hypothesis tests for comparing a population variance, σ_1^2, with another population variance, σ_2^2, are given by the following:

H_0: $\sigma_1^2 = \sigma_2^2$	H_0: $\sigma_1^2 \leq \sigma_2^2$	H_0: $\sigma_1^2 \geq \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$	H_1: $\sigma_1^2 > \sigma_2^2$	H_1: $\sigma_1^2 < \sigma_2^2$

The null hypothesis is denoted by H_0 and the alternative hypothesis is denoted by H_1. The shape of the F distribution is non-symmetrical and will depend on the number of degrees of freedom associated with S_1^2 and S_2^2. These quantities are represented by ν_1 and ν_2 respectively. (ν_1 is the DF in the numerator.)

The F statistic is the ratio of two sample variances (two chi square distributions) and is given by the formula:

$$F = \frac{s_1^2}{s_2^2}$$

Where s_1^2 and s_2^2 are sample variances.

Since the identification of the sample variances is arbitrary, it is customary to designate the larger sample variance as s_1^2 and place it in the numerator.

Paired-Comparison Hypotheses Tests (Continued)

F Test (Continued)

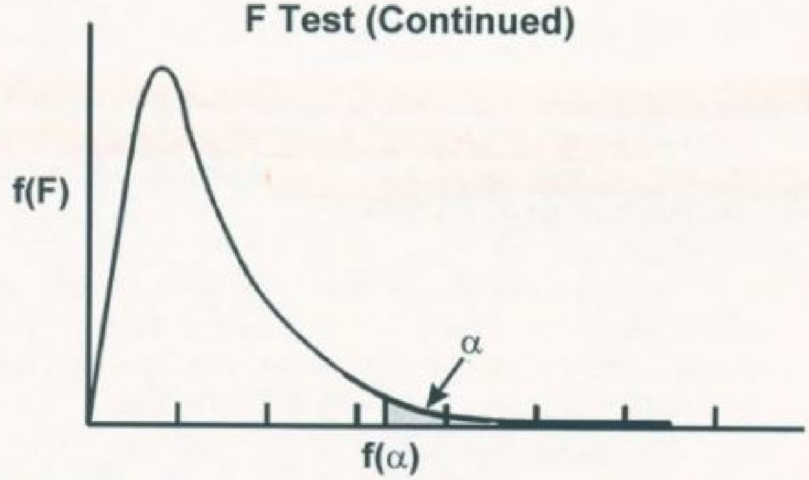


Figure 8.27 An F Distribution for Two Normal Samples Both with DF = 10

There are numerous F Table formats based on α values of 0.10, 0.05, 0.025, 0.01, etc. Listed below is a partial F Table for α = 0.05.

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

Table 8.28 F Critical Values (α=0.05)

Note: The above critical values for F may be used for a one-tailed test (α=0.05, 95% confidence) or a two-tailed test (α/2=0.05, 90% confidence).

Paired-Comparison Hypotheses Tests (Continued)

F Test (Continued)

Example 8.22: A materials laboratory is studying the effect of aging on a product. They want to know if there is an improvement in consistency of strength after aging for one year (assume a 95% confidence level). The data obtained is listed below:

	At Start	One Year Later
No. of tests	9	7
Product standard deviation (psi)	900	300

Solution: H_0: $\sigma_1^2 \leq \sigma_2^2$ H_1: $\sigma_1^2 > \sigma_2^2$ and $DF_1 = 8$ $DF_2 = 6$

The concerned is with an improvement in variation; therefore, use a one-tail test with the entire α risk in the right tail. From the prior F Table, one sees that the critical value of F is 4.15. The null hypothesis rejection area is equal to or greater than 4.15.

$$F = \frac{S_1^2}{S_2^2} = \frac{(900)^2}{(300)^2} = 9$$

Conclusion: Since the calculated F value is in the critical region, the null hypothesis is rejected. There is sufficient evidence to indicate a reduced variance and more consistency of strength after aging for one year.

Example 8.23: Two samples consisting of 10 and 8 measurements were observed to have variances of $s_1^2 = 7.14$ and $s_2^2 = 3.21$, respectively. Do the sample variances present sufficient evidence to indicate that the population variances are different at a 90% confidence level?

Solution: H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 \neq \sigma_2^2$ $DF_1 = 9$ $DF_2 = 7$ $\alpha/2 = 0.05$

The test statistic is:

$$F = \frac{S_1^2}{S_2^2} = \frac{7.14}{3.21} = 2.22$$

From Table 8.28, the right tail critical value of F = 3.68. The left tail critical value of F is found using the equation on page VII - 39 where F = 1/(3.29) = 0.304, where the degrees of freedom terms are 7 and 9 when finding the table value. Since the calculated F value of 2.22 is between the critical values of 0.304 and 3.68, one fails to reject the null hypothesis and cannot conclude that the population variances are different at a 90% confidence level.

Summary of Inference Tests

TYPE	TEST STATISTIC	DF	APPLICATION
Z	$Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$	N.A.	Single sample mean. Standard deviation of population is known.
t Test	$t = \dfrac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}}$	n - 1	Single sample mean. Standard deviation of population unknown.
2 Mean Equal Variance t Test	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{{}^*S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$n_1 + n_2 - 2$	2 sample means. Variances are unknown, but considered equal. $^*S_p = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$
2 Mean Unequal Variance t Test	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	*	2 sample means. Variances are unknown, but considered unequal. DF is determined from the Welch-Satterthwaite
Paired t Test	$t = \dfrac{\bar{d}}{s_d/\sqrt{n}}$	n - 1	2 sample means. Data is taken in pairs. A different d is calculated for each pair.
X^2 σ^2 Known	$X^2 = \dfrac{(n-1)S^2}{\sigma^2}$	n - 1	Tests sample variance against known variance.
X^2	$X^2 = \sum \dfrac{(O - E)^2}{E}$	(r-1)(c-1)	Compares observed and expected frequencies of test outcomes.
F	$F = \dfrac{S_1^2}{S_2^2}$	$n_1 - 1$ $n_2 - 1$	Tests if two sample variances are equal.

Table 8.29 Comparison Summary of Inference Tests

The student may wish to refer to *Juran's Quality Handbook* (Juran, 1999)[14] or *Business Statistics* (Triola, 1994)[24] for a more comprehensive summary of inference tests.

Analysis of Variance Introduction

Earlier, techniques were presented for estimating and testing the values of a single population mean or the difference between two means (Z test and student's t test). In many investigations, such as experimental trials, it is necessary to compare three or more population means simultaneously. There are underlying assumptions in this analysis of variance of means: the variance is the same for all factor treatments or levels, the individual measurements within each treatment are normally distributed and the error term is considered a normally and independently distributed random effect.

With analysis of variance, the variations in response measurement are partitioned into components that reflect the effects of one or more independent variables. The variability of a set of measurements is proportional to the sum of squares of deviations used to calculate the variance:

$$\Sigma(X - \bar{X})^2$$

Analysis of variance partitions the sum of squares of deviations of individual measurements from the grand mean (called the total sum of squares) into parts: the sum of squares of treatment means plus a remainder which is termed the experimental or random error. When an experimental variable is highly related to the response, its part of the total sum of the squares will be highly inflated. This condition is confirmed by comparing the variable sum of squares with that of the random error sum of squares using an F test.

A Comparison of Three or More Means

An analysis of variance to detect a difference in three or more population means first requires obtaining some summary statistics for calculating variance of a set of data as shown below:

Where:

ΣX^2 is called the crude sum of squares

$(\Sigma X)^2 / N$ is the CM (correction for the mean), or CF (correction factor)

$\Sigma X^2 - (\Sigma X)^2 / N$ is termed SS (total sum of squares, or corrected SS).

$$\frac{\Sigma X^2 - (\Sigma X)^2 / N}{N - 1} = \sigma^2 (\text{variance}) = \frac{\text{Total Sum of Squares}}{\text{Total DF (degrees of freedom)}}$$

A Comparison of Three or More Means (Continued)

One-Way ANOVA

In the one-way ANOVA, the total variation in the data has two parts: the variation among treatment means and the variation within treatments.

Let the ANOVA grand average = $\Sigma X/N$ = GM. The total SS (Total SS) is then:

Total SS = $\Sigma(X_i - GM)^2$ Where X_i is any individual measurement.

Total SS = SST + SSE Where SST = treatment sum of squares and SSE is the experimental error sum of squares.

$SST = \sum n_t \left(\bar{X}_t - GM\right)^2$ = Sum of the squared deviations of each treatment average from the grand average or grand mean.

$SSE = \sum \left(X_t - \bar{X}_t\right)^2$ = Sum of the squared deviations of each individual observation within a treatment from the treatment average.

For the ANOVA calculations:

$$\sum(TCM) = \sum \frac{\text{Each treatment total squared}}{\text{No. obs. in that treatment}}$$ (Treatment total CM)

SST = Σ(TCM) - CM

SSE = Total SS - SST (Always obtained by difference)

Total DF = N - 1 (Total Degrees of Freedom)

TDF = t - 1 (Treatment DF = Number of treatments minus 1)

EDF = (N - 1) - (t - 1) = N - t (Error DF, always obtained by difference)

$$MST = \frac{SST}{TDF} = \frac{SST}{t-1}$$ (Mean Square Treatments)

$$MSE = \frac{SSE}{EDF} = \frac{SSE}{N-t}$$ (Mean Square Error)

A Comparison of Three or More Means (Continued)

One-Way ANOVA (Continued)

To test the null hypothesis:

$H_O: \mu_1 = \mu_2 = ... = \mu_t$ $\quad$ H_1: At least one mean different

$F = \frac{MST}{MSE}$ $\quad$ When $F > F_\alpha$, reject H_O

Example 8.24: As an example of a comparison of three means, consider a single factor experiment: The following coded results were obtained from a single factor randomized experiment, in which the outputs of three machines were compared. Determine if there is a significant difference in the results ($\alpha = 0.05$).

Machines	Data	Sum	n	Avg	TCM = $(\text{Sum})^2/n$	ΣX^2
A	5, 7, 6, 7, 6	31	5	6.2	192.2	195
B	2, 0, 1, -2, 2	3	5	0.6	1.8	13
C	1, 0, -2, -3, 0	-4	5	-0.8	3.2	14
Total		30	15		197.2	222

$\Sigma X = 30 \quad N = 15 \quad \text{Total DF} = N - 1 = 15 - 1 = 14$

$GM = \sum X/N = 30/N = 2.0 \qquad \sum X^2 = 222$

$$CM = \frac{\left(\sum X\right)^2}{N} = \frac{(30)^2}{15} = 60$$

$\text{Total SS} = \sum X^2 - CM = 222 - 60 = 162$

$\sum(TCM) = 197.2$

$SST = \sum(TCM) - CM = 197.2 - 60 = 137.2$ and

$SST = \sum n_t\left(X_t - GM\right)^2 = 5(6.2 - 2)^2 + 5(0.6 - 2)^2 + 5(-0.8 - 2)^2$

$SST = 88.2 + 9.8 + 39.2 = 137.2$

$SSE = \text{Total SS} - SST = 162 - 137.2 = 24.8$

A Comparison of Three or More Means (Continued)

One-Way ANOVA (Continued)

Example 8.24 (continued):

The completed ANOVA table is:

Source (of variation)	SS	DF	Mean Square	F	F_{α, v_1, v_2}
Machines	137.2	2	68.6	33.2	$F_{0.05, 2, 12} = 3.89$
Error	24.8	12	2.067		
Total	162	14			

$\sigma_e = \sqrt{2.07} = 1.44$

Since the computed value of F (33.2) exceeds the critical value of F, the null hypothesis is rejected. Thus, there is evidence that a real difference exists among the machine means.

σ_e is the pooled standard deviation of within treatments variation. It can also be considered the process capability sigma of individual measurements. It is the variation within measurements which would still remain if the difference among treatment means were eliminated.

Two-Way ANOVA

It will be seen that the two-way analysis procedure is an extension of the patterns described in the one-way analysis. Recall that a one-way ANOVA has two components of variance: Treatments and experimental error (may be referred to as columns and error or rows and error). In the two-way ANOVA there are three components of variance: Factor A treatments, Factor B treatments, and experimental error (may be referred to as columns, rows, and error).

2 Factor, Two-Way ANOVA Experiment

Example 8.25: In a hypothetical example, three different CSSBB study materials were taught by two different instructors to three different students with the following results. The responses are examination results as a percentage.

	Materials						
Instruct	1	2	3	Sum	n	Avg	RowSq
1	88 84 92	68 62 72	76 70 78	690	9	76.667	52900
2	60 68 58	44 50 52	60 52 56	500	9	55.556	27778
Sum	450	348	392	1190			80678
n	6	6	6		18		
Avg	75	58	65.33				
ColSq	33750	20184	25610.7	79544.7			
ΣX^2	34852	20792	26200	81844			

The null hypothesis: instructor and study material means do not differ.

$\Sigma X = 1190$ $N = 18$ Total DF = 17 $GM = \Sigma X/N = 1190/18 = 66.11$

$CM = (\Sigma X)^2/N = (1190^2)/18 = 78672.22$ $\Sigma X^2 = 81844$

Total SS = $\Sigma X^2 - CM = 81844 - 78672.22 = 3171.78$

ColSq = column total squared and divided by the no. of observations in the column

RowSq = row total squared and divided by the no. of observations in the row

2 Factor, Two-Way ANOVA Experiment

Example 8.25 (continued):

SSCol = ΣColSq - CM = 79544.67 - 78672.22 = 872.44

SSRow = ΣRowSq - CM = 80677.78 - 78672.22 = 2005.56

SSE = Total SS - SSCol - SSRow = 3171.78 - 872.44 - 2005.56 = 293.78

The next step is to construct the ANOVA table.

Source	SS	DF	MS	F	$F_{0.05,\gamma_1,\gamma_2}$
Columns (materials)	872.44	2	436.22	20.8	$F_{0.05,2,14} = 3.74$
Rows (instructor)	2005.56	1	2005.56	95.6	$F_{0.05,1,14} = 4.60$
Error	293.78	14	20.98		
Total		17			

SIGe = $\sqrt{20.98}$ = 4.58 SIG total = $\sqrt{\text{Total SS}/(N - 1)}$ = 13.66

If no interaction: Col DF = Col - 1 = 3 - 1 = 2 Row DF = Row - 1 = 2 - 1 = 1
Error DF = Total DF - Col DF - Row DF = 17 - 2 - 1 = 14

Col F = MSCol/MSE = 436.22/20.98 = 20.79. This is larger than critical F = 3.74. Therefore, the null hypothesis of equal material means is rejected.

Row F = MSRow/MSE = 2005.56/20.98 = 95.59. This is larger than critical F = 4.60. Therefore, the null hypothesis of equal instructor means is rejected.

The difference between total sigma (13.66) and error sigma (4.58) is due to the significant difference in instructor means and material means. If the instructor and study material differences were only due to chance cause, the sigma variation in the data would be equal to SIGe, the square root of the mean square error.

2 Factor ANOVA Experiment with Interaction

It should be noted, in the materials/instructor example above, the data was listed in six cells. That is, six experimental combinations. There were also 3 replications (students) in each cell (k = 3). When k is greater than 1 in a 2 factor ANOVA, there is the opportunity to analyze for a possible interaction between the two factors.

Example 8.25 (continued): Examine the previous data for interaction effects. A similar analysis pattern is noted here. The data in each cell is summed, and that total is divided by the number of observations in that cell.

$$\text{CellSq} = \frac{(\text{SumCell})^2}{k} \qquad \text{InterSq} = \sum(\text{CellSq})$$

$$\text{SSInter} = \text{InterSq} - \text{CM} - \text{SSCol} - \text{SSRow}$$

For the sum of squares interaction (SS Inter), it is not enough to just subtract the correction for the mean (CM) as was done to determine the main effects of SSCol and SSRow. This is because the data in replicated cells is affected by the treatment levels of the two factors of which it is a part as well as a possible interaction effect. To net out the interaction effect, it is necessary to also subtract the sum of squares column and row factors previously calculated. The cell-by-cell calculations are shown below.

	Materials		
Instructor	1	2	3
1	264 Avg=88	202 Avg=67.33	224 Avg=74.67
2	186 Avg=62	146 Avg=48.67	168 Avg=40.33

$$\sum(\text{CellSq}) = \frac{(264)^2}{3} + \frac{(202)^2}{3} + \frac{(224)^2}{3} + \frac{(186)^2}{3} + \frac{(146)^2}{3} + \frac{(168)^2}{3}$$

$$\sum(\text{CellSq}) = 23232 + 13601.33 + 16725.33 + 11532 + 7105.33 + 9408$$

$$\sum(\text{CellSq}) = 81604$$

2 Factor ANOVA with Interaction (Continued)

Example 8.25 (continued):

SSInter = 81604 - 78672.22 - 872.44 - 2005.56 = 53.78

SSError = TotSS - SSCol - SSRow - SSInter

SSError = 3171.78 - 872.44 - 2005.56 - 53.78 = 240

The null hypothesis for the interaction effect is that there is no interaction. See the revised ANOVA table below:

ANOVA TABLE

Source	SS	DF	MS	F	$F_{0.05,v_1,v_2}$
Columns (materials)	872.44	2	436.22	21.81	$F_{0.05,2,12}$ = 3.89
Rows (instructor)	2005.6	1	2005.56	100.3	$F_{0.05,1,12}$ = 4.75
Interaction (Row/Col)	53.78	2	26.89	1.34	$F_{0.05,2,12}$ = 3.89
Error	240	12	20		
		17			

SIGe = $\sqrt{20}$ = 4.47 SIG total = $\sqrt{\text{Total SS}/(N-1)}$ = 13.66

With interaction: Replications per cell = k = 3
Col DF = Col - 1 = 3 - 1 = 2 Row DF = Row - 1 = 2 - 1 = 1
Inter DF = (Col - 1)(Row - 1) = (3 - 1)(2 - 1) = 2
Error DF = Total DF - Col DF - Row DF - Inter DF = 17 - 2 - 1 - 2 = 12

The interaction calculated F (1.34) is less than critical F (3.89). The null hypothesis of no interaction is not rejected. There is an advantage in analyzing for possible interaction if the opportunity exists. The more effects which are significant, the greater the amount of total variation which is explained and the smaller the MS error (unexplained variation). As the MS error is the divisor in the F ratio, a smaller MS error increases the sensitivity of testing effects.

Components of Variance

The analysis of variance can be extended with a determination of the COV (components of variance). The COV table uses the MS (mean square), F, and F(alpha) columns from the previous ANOVA TABLE and adds columns for EMS (expected mean square), variance, adjusted variance and percent contribution to design data variation. The model for the ANOVA is:

$$X_{ijk} = \mu + M_i + I_j + MI_{ij} + \varepsilon_{ijk}$$

The model states that any measurement (X) represents the combined effect of the population mean (μ), the different materials (M), the different instructors (I), the materials/instructor interaction (MI), and the experimental error (ε). Where: I represents materials at 3 levels, j represents instructors at 2 levels, k represents 3 replications per cell.

Example 8.25 (continued):

COV TABLE

	MS	F	F(α)	EMS	VAR	ADJ VAR	% CONTR
Col	436.22	21.81	3.89	$\sigma_e^2 + 6\sigma_M^2$	69.37	69.37	22.21
Row	2005.6	100.3	4.75	$\sigma_e^2 + 9\sigma_I^2$	220.62	220.62	70.65
Inter	26.89	1.34	3.89	$\sigma_e^2 + 3\sigma_{MI}^2$	2.3	2.3	0.74
Error	20			σ_e^2	20	20	6.4
					Totals	312.39	100

The variance coefficients are equal to the number of values used in calculating the respective MS. Materials coef = k x Row = 3 x 2 = 6, instructors coef = k x Col = 3 x 3 = 9 Interaction coef = k = 3. The general variance equation is given by:

Effect Variance = (MS Effect - MS Error)/(Variance Coefficient)

M Var = (436.22 - 20)/6 = 69.37 I Var = (2005.56 - 20)/9 = 220.62

MI Var = (26.89 - 20)/3 = 2.30 Error Var = 20

Material differences are significant and account for 22.21% of the variation in the data. Instructor differences are significant and account for 70.65% of variation in the data. The material/instructor interaction is not significant and shows as a negligible contribution. Experimental error accounts for only 6.40% of the total variation. The reason for the adjusted variance column is that variance calculations are negative when the mean square effect is less than the mean square error. Negative mean squares are considered to have a value of 0. Knowing the percent contribution aids in establishing priorities when taking improvement actions.

ANOVA Table for an A x B Factorial Experiment

In a factorial experiment involving factor A at a levels and factor B at b levels, the total sum of squares can be partitioned into:

Total SS = SS(A) + SS(B) + SS(AB) + SSE

ANOVA Table for an A x B Factorial Experiment			
Source	DF	SS	MS
Factor A Factor B Interaction AB Error	(a-1) (b-1) (a-1)(b-1) (N-ab)	SS(A) SS(B) SS(AB) SSE	SS(A)/(a-1) SS(B)/(b-1) SS(AB)/(a-1)(b-1) SSE/(N-ab)
Total	(N-1)	Total SS	

Table 8.30 ANOVA for an A x B Factorial Experiment

ANOVA Table for a Randomized Block Design

The randomized block design implies the presence of two independent variables, blocks and treatments. The total sum of squares of the response measurements can be partitioned into three parts, the sum of the squares for the blocks, treatments, and error. The analysis of a randomized block design is of less complexity than an A x B factorial experiment.

ANOVA Table for a Randomized Block Design			
Source	DF	SS	MS
Blocks Treatments Error	b-1 t-1 (b-1)(t-1)	SSB SST SSE	MSB=SSB/(b-1) MST=SST/(t-1) MSE=SSE/(b-1)(t-1)
Total	bt-1	Total SS	

Table 8.31 ANOVA for a Randomized Block Design

Goodness-of-Fit Tests

GOF (goodness-of-fit) tests are part of a class of procedures that are structured in cells. In each cell there is an observed frequency, (F_o). From the nature of the problem, one either knows the expected or theoretical frequency, (F_e) or can calculate it. Chi square (X^2) is then summed across all cells according to the formula:

$$X^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

The calculated chi square is then compared to the chi square critical value for the following appropriate degrees of freedom:

Goodness-of-fit distribution	Degrees of freedom (DF)
Normal	No. of cells - 3
Poisson	No. of cells - 2
Binomial	No. of cells - 2
Uniform	No. of cells - 1

Uniform Distribution (GOF)

Example 8.26: Is a game die balanced? The null hypothesis, H_0, states the die is honest and balanced. When a die is rolled, the expectation is that each side should come up an equal number of times. It is obvious there will be random departures from this theoretical expectation if the die is honest. A die was tossed 48 times with the following results:

Spots	F_e	F_o	$(F_e - F_o)^2 / F_e$
1	8	12	2
2	8	7	0.125
3	8	2	4.5
4	8	7	0.125
5	8	12	2
6	8	8	0
Total =	48	48	8.75

The calculated chi square is 8.75. The critical chi square $X^2_{0.05, 5} = 11.07$. The calculated chi square does not exceed critical chi square. Therefore, the hypothesis of an honest die cannot be rejected. The random departures from theoretical expectation could well be explained by chance cause.

Normal Distribution (GOF)

Example 8.27: The following data (105 observations) is taken from an $\bar{X}$ - R chart (closure removal torques) presented in Section X. There is sufficient data for ten cells. The alternative would be six cells which is too few. Twelve integer cells fit the range of the data. The null hypothesis: the data was obtained from a normal distribution.

A	B	C	D	E	F	G	H	I	J	K
(CB) cell Boundary	Cell Middle	Obs Freq Fo	CB - $\bar{X}$	Z= (CB-$\bar{X}$)/ SD	Cum Prob	Cell Prob	Theo Freq Fe	Theo Freq Pooled	Obs Freq Pooled	Chi Square
22.5			7.1	4.4099	1.0000					
	22	1				0	0.01			
21.5			6.1	3.7888	0.9999					
	21	0				0	0.073			
20.5			5.1	3.1677	0.9992					
	20	0				0	0.4904			
19.5			4.1	2.5466	0.9946					
	19	4				0.0220	2.2729			
18.5			3.1	1.9255	0.9729					
	18	8				0.0690	7.2421	10.086	13	0.842
17.5			2.1	1.3043	0.9039					
	17	19				0.1512	15.873	15.873	19	0.616
16.5			1.1	0.6832	0.7528					
	16	21				0.228	23.941	23.941	21	0.361
15.5			0.1	0.062	0.5248					
	15	24				0.2367	24.852	24.852	24	0.029
14.5			-0.9	-0.559	0.2881					
	14	11				0.1691	17.756	17.756	11	2.571
13.5			-1.9	-1.18	0.1190					
	13	6				0.0831	8.73	12.479	17	1.638
12.5			-2.9	-1.801	0.0358					
	12	6				0.0281	2.9528			
11.5			-3.9	-2.422	0.0077					
	11	3				0.0100	0.6868			
10.5			-4.9	-3.044	0					
	10	2				0	0.1098			
9.5			-5.9	-3.665	0					
									X^2_{cal} =	6.057

$\bar{X}$ = 15.4, sigma = 1.61, number of effective cells = 6, DF = 3 and $X^2_{0.05,\,3}$ = 7.81

Normal Distribution (GOF) (Continued)

Example 8.27 (continued): One degree of freedom is lost because $\bar{X}$ estimates μ. A second degree of freedom is lost because SD estimates sigma. A third degree of freedom is lost because sample N represents the population.

Col A: The cell boundaries are one half unit from the cell midpoint.

Col B: The cell middle values are integers.

Col C: The observed frequencies in each cell are Fo.

Col D: Distances from $\bar{X}$ are measured from cell boundaries.

Col E: Distances from $\bar{X}$ are divided by SD to transform distances into Z units.

Col F: Z units are converted into cumulative normal distribution probabilities.

Col G: The theoretical probability in each cell is obtained by taking the difference between cumulative probabilities in Column F. The top cell theoretical probability boundary is 1.0000.

Col H: The theoretical frequency in each cell is the product of N and Column G.

Col I: Each cell is required to have a theoretical frequency equal to or greater than four. Therefore, the top four cells must be added to the cell whose midpoint is 18. The bottom three cells must be added to the cell whose midpoint is 13. Thus, there are six effective cells, all of which have a theoretical frequency equal to or greater than four.

Col J: The observed frequency cells must be pooled to match the theoretical frequency cells. It does not matter if the observed frequencies are less than four.

Col K: The contributions to chi square are obtained by squaring the difference between Column I and Column J and dividing by Column I.

$$\frac{(\text{Column I - Column J})^2}{\text{Column I}}$$

Conclusion: Since the calculated chi square, 6.057, is less than critical chi square, 7.81, we fail to reject the null hypothesis of normality, and therefore, conclude the data is from a normal distribution.

Poisson Distribution (GOF)

Example 8.28: The bead drum is an attribute variable, random sample generating device, which was used to obtain the following data. In this exercise red beads represent defects. Seventy-five constant size samples were obtained. The goodness-of-fit test is analyzed based on sample statistics. The null hypothesis is that the bead drum samples represent a Poisson distribution.

A	B	C	D	E	F	G	H
No. of Defects c	Prob (c)	Theo Freq Fe	Obs Freq Fo	Fe Pooled	Fo Pooled	$\frac{(F_e - F_O)^2}{F_e}$	(c)(Fo)
0	0.0277	2.0768	1				0
1	0.0993	7.4487	6	9.5255	7	0.6696	6
2	0.1781	13.3581	11	13.3581	11	0.4163	22
3	0.2129	15.9703	22	15.9703	22	2.2765	66
4	0.1909	14.3201	14	14.3201	14	0.0072	56
5	0.137	10.2722	12	10.2722	12	0.2906	60
6	0.0819	6.1405	5	6.1405	5	0.2118	30
7	0.042	3.1463	3	5.406	4	0.3657	21
8	0.0188	1.4106	1				8
9	0.0075	0.5621					
10	0.0027	0.2016					
11	0.0009	0.0657					
12	0.0003	0.0196					
Totals	0.9999	75	75	75	75	$X^2 = 4.238$	269

N = 75 Sample Avg = 269/75 = 3.59 DF = 7 - 2 = 5 $X^2_{0.05,\,5} = 11.07$

One degree of freedom is lost because $\bar{c}$ (sample average = 3.59) estimates μ. A second degree of freedom is lost because N (number of samples) estimates the population.

Col A: Values of c which matched the actual distribution of sample defects found.

Col B: The probability that c defects would occur given the average value of the samples.

Col C: The theoretical number of defects that would occur (N x Col B).

Col D: The observed frequency of each number of defects.

Poisson Distribution (GOF) (Continued)

Example 8.28 (continued):

Col E: The required minimum frequency of four for each effective cell resulted in pooling at both tails of the theoretical Poisson distribution.

Col F: The observed frequency distribution of defects must also be pooled to match the effective theoretical distribution.

Col G: The contributions to chi square are obtained from squaring the difference between Fe and Fo and dividing the result by Fe.

Col H: Total defects found result from the product of number of defects and observed frequency.

Since the calculated chi square of 4.238 is less than the critical chi square value of 11.07 at the 95% confidence level, we fail to reject the null hypothesis that the bead drum samples represent a Poisson distribution.

Binomial Distribution (GOF)

Example 8.29: The null hypothesis states that the following industrial sample data comes from a binomial population of defectives (N = 80). In this case, we will estimate the probability of a defective from the sample data, p = 0.025625.

A	B	C	D	E	F	G	H
Defectives d	Prob (d)	Theo Freq Fe	Obs Freq Fo	Theo Frequency Pooled	Observed Frequency Pooled	Chi Square	(d)(Fo)
0	0.1253	10.03	10	10.03	10	0	0
1	0.2637	21.1	30	21.1	30	3.76	30
2	0.2739	21.92	17	21.92	17	1.1	34
3	0.1873	14.98	8	14.98	8	3.26	24
4	0.0948	7.59	6	7.59	6	0.33	24
5	0.0379	3.03	4	4.39	9	4.85	20
6	0.0125	1	4				24
7	0.0035	0.28	0				0
8	0.0008	0.07	1				8
9	0.0002	0.01					
Totals	1	80	80	80	80	X^2=13.30	164

Average ($\bar{d}$) = 164/80 = 2.05, p = 2.05/80 = 0.025625 DF = 6 - 2 = 4 $X^2_{0.05,\,4}$ = 9.49

Binomial Distribution (GOF) (Continued)

Example 8.29 (continued): One degree of freedom is lost because the total sample frequency represents the population. A second degree of freedom is lost because $\bar{d}$ is used to estimate μ.

Col A: The range of defectives matching the observed sample data.

Col B: The probability of observed cell defective count given sample size N and $\bar{d}$.

Col C: The expected theoretical frequency (cell probability)(N).

Col D: The observed cell frequency count from the 80 samples.

Col E: Theoretical frequency with cells pooled to meet n = 4 minimum.

Col F: Observed cell frequency pooled to match theoretical frequency pooled cells.

Col G: Contributions to chi square $(Fe - Fo)^2/Fe$.

Col H: The count of defectives by cell (d)(Fo).

The calculated chi square = 13.30. The critical chi square = 9.49. Since the calculated value is greater than the critical value, the null hypothesis that the sample data represents the binomial distribution is rejected at the 95% confidence level.

Contingency Tables

A two-way classification table (rows and columns) containing original frequencies can be analyzed to determine whether the two variables (classifications) are independent or have significant association. R. A. Fisher determined that when the marginal totals (of rows and columns) are analyzed in a certain way, that the chi square procedure will test whether there is dependency between the two classifications. In addition, a contingency coefficient (correlation) can be calculated. If the chi square test shows a significant dependency, the contingency coefficient shows the strength of the correlation.

It often happens that results obtained in samples do not always agree exactly with the theoretical expected results according to rules of probability. A measure of the difference found between observed and expected frequencies is supplied by the statistic chi square, X^2, where:

$$X^2 = \sum_{n=1}^{k} \frac{(O_n - E_n)^2}{E_n} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \dots + \frac{(O_k - E_k)^2}{E_k}$$

If $X^2 = 0$, the observed and theoretical frequencies agree exactly. If $X^2 > 0$, they do not agree exactly. The larger the value of X^2, the greater the discrepancy between observed and theoretical frequencies. The chi square distribution is an appropriate reference distribution for critical values when the expected frequencies are at least equal to 5.

Example 8.30: The calculation for the E (expected or theoretical) frequency will be demonstrated in the following example. Five hospitals tried a new drug to alleviate the symptoms of emphysema. The results were classified at three levels: no change, slight improvement, marked improvement. The percentage matrix is shown in Table 8.32.

	Hospital					Average Row Percentage
Results	A	B	C	D	E	
No change	28.3	16.7	9.4	14.7	53	24.4
Slight improvement	37.7	30.6	45.9	44.9	36.1	39
Marked improvement	34	52.8	44.7	40.4	10.8	36.5

Table 8.32 Percentage Matrix

Contingency Tables (Continued)

Example 8.30 (continued): While the results expressed as percentages do suggest differences among hospitals, ratios presented as percentages can be misleading. A proper analysis requires that original data be considered as frequency counts. Table 8.33 lists the original data on which the percentages are based.

	Hospital					Actual Frequencies
Results	A	B	C	D	E	
No change	15	6	8	20	44	93
Slight improvement	20	11	39	61	30	161
Marked improvement	18	19	38	55	9	139
Total Frequency	53	36	85	136	83	393

Table 8.33 Original Data

The calculation of expected, or theoretical, frequencies is based on the marginal totals. The marginal totals for the frequency data are the column totals, the row totals, and the grand total. The null hypothesis is that all hospitals have the same proportions over the three levels of classifications. To calculate the expected frequencies for each of the 15 cells under the null hypothesis requires the manipulation of the marginal totals as illustrated by the following calculation for one cell. Consider the count of 15 for Hospital A/no change cell. The expected value, E, is:

$$E = \frac{\text{column total x row total}}{\text{grand total}} = \frac{(53)(93)}{393} = 12.5$$

The same procedure repeated for the other 14 cells yields Table 8.34.

12.5	8.5	20.1	32.2	19.6
21.7	14.7	34.8	55.7	34
18.7	12.7	30.1	48.1	29.4

Table 8.34 Theoretical Frequencies

Contingency Tables (Continued)

Example 8.30 (continued): Each of these 15 cells makes a contribution to chi square (X^2). For the same selected (illustrative) cell, the contribution is:

$$\frac{(O - E)^2}{E} = \frac{(\text{Observed frequency - Expected frequency})^2}{\text{Expected frequency}}$$

$$\frac{(O - E)^2}{E} = \frac{(15 - 12.5)^2}{12.5} = 0.48$$

Chi Square = $\sum \frac{(O - E)^2}{E}$ over all cells. See the values in Table 8.35.

0.48	0.74	7.3	4.61	30.21
0.14	0.95	0.5	0.5	0.47
0.03	3.08	2.1	0.99	14.12
			Chi Sq =	66.22

Table 8.35 Contributions to Chi Square

Assume alpha to be 0.01. The degrees of freedom for contingency tables is:

d.f. = (rows - 1) x (columns - 1).

For this example: d.f. = (5 - 1) x (3 - 1) = 8

The critical chi square: $X^2 = X^2_{0.01,\,8} = 20.09$

The calculated chi square is larger than critical chi square. Therefore, one rejects the null hypothesis of hospital equality of results. The alternative hypothesis is that hospitals differ.

Coefficient of Contingency (C)

The degree of relationship, association or dependence of the classifications in a contingency table is given by:

$$C = \sqrt{\frac{X^2}{X^2 + N}}$$

Where N equals the grand frequency total.

Example 8.30 (continued): The contingency coefficient is:

$$C = \sqrt{\frac{X^2}{X^2 + N}} = \sqrt{\frac{66.22}{66.22 + 393}} = 0.38$$

The maximum value of C is never greater than 1.0, and is dependent on the total number of rows and columns. For the example data, the maximum coefficient of contingency is:

$$\text{Max } C = \sqrt{\frac{k - 1}{k}} = \sqrt{\frac{3 - 1}{3}} = 0.816$$

Where: k = min of (r, c) and r = rows, c = columns

There is a Yates correction for continuity test that can be performed when the contingency table has exactly two columns and two rows. That is, the degrees of freedom is equal to 1. The application of this correction is found in a number of statistical sources, but is outside the scope of this Primer.

Correlation of Attributes

Contingency table classifications often describe characteristics of objects or individuals. Thus, they are often referred to as attributes and the degree of dependence, association, or relationship is called correlation of attributes. For (k = r = c) tables, the correlation coefficient, φ, is defined as:

$$\varphi = \sqrt{\frac{X^2}{N(k - 1)}}$$

The value of φ falls between 0 and 1. If the calculated value of chi square is significant, then φ is significant. In the above example, rows and columns are not equal and the correlation calculation is not applied.

Parametric vs. Nonparametric Tests

Parametric implies that a distribution is assumed for the population. Often, an assumption is made when performing a hypothesis test that the data is a sample from a certain distribution, commonly the normal distribution. Nonparametric implies that there is no assumption of a specific distribution for the population.

An advantage of a parametric test is that if the assumptions hold, the power, or the probability of rejecting H_0, when it is false, is higher than the power of a corresponding nonparametric test with equal sample sizes. An advantage of nonparametric tests is that the test results are more robust against violation of the assumptions. Therefore, if assumptions are violated for a test based upon a parametric model, the conclusions based on parametric test p-values may be more misleading than conclusions, based upon nonparametric test p-values.

Nonparametric Techniques

Nonparametric techniques of hypothesis testing are applicable for many quality engineering problems and projects. The nonparametric tests are often called "distribution-free" since they make no assumption regarding the population distribution. Nonparametric tests may be applied ranking tests in which data is not specific in any continuous sense, but are simply ranks.

Three nonparametric techniques will be described with examples:

- The Kruskal Wallis One Way ANOVA
- The Mann Whitney U Test
- Wilcoxon-Mann-Whitney Rank Sum Test

Kruskal-Wallis One-Way Analysis of Variance by Ranks

This is a test of independent samples. The measurements may be continuous data, but the underlying distribution is either unknown, or known to be non-normal. In either case, the data can be ranked and analyzed without the constraint of having to assume a known population distribution.

Example 8.31: Three different plants manufactured the same garment style. Variation in garment length was a customer concern. Length was measured to the nearest 1/4". Within each plant, only four measurement increment values were obtained. This lack of measurement sensitivity indicated that ranking the data was preferred to assuming normality. The null hypothesis is that the population medians are the same. H_o: $M_1 = M_2 = ... = M_n$. The following table shows data coded as deviations from a common reference value.

Original Data Measurements (Coded)

Plant A	Plant B	Plant C
0.25	0.50	0.25
0.25	1.00	1.00
0.50	0.25	1.00
1.00	0.75	0.75
0.50	0.25	1.00
0.50	0.25	0.50
0.25		1.00
0.75		

For simplicity and convenience, the coded data can be further coded as integers.

1	2	1
1	4	4
2	1	4
4	3	3
2	1	4
2	1	2
1		4
3		

The next step is to construct a combined sample, rank the combined data while retaining plant identity, and reconstitute the three plant sample sets with ranks replacing the original data. Tied ranks are replaced by the average value of the ties.

Kruskal-Wallis (Continued)

Example 8.31 (continued):

COMBINED SAMPLES		RANK COMBINED SAMPLES		RANK IF THERE WERE NO TIES	RANKING WITH TIES	
Plant	X	PLANT	X		PLANT	TIED RANK
A	1	A	1	1	A	4
A	1	A	1	2	A	4
A	2	A	1	3	A	4
A	4	B	1	4	B	4
A	2	B	1	5	B	4
A	2	B	1	6	B	4
A	1	C	1	7	C	4
A	3	A	2	8	A	10
B	2	A	2	9	A	10
B	4	A	2	10	A	10
B	1	B	2	11	B	10
B	3	C	2	12	C	10
B	1	A	3	13	A	14
B	1	B	3	14	B	14
C	1	C	3	15	C	14
C	4	A	4	16	A	18.5
C	4	B	4	17	B	18.5
C	3	C	4	18	C	18.5
C	4	C	4	19	C	18.5
C	2	C	4	20	C	18.5
C	4	C	4	21	C	18.5

There were seven coded values tied at 4. They would have been ranks 1 through 7.

The average of ranks 1 - 7 is 4. All coded measurement values of 1 received the average rank of 4. In a similar fashion, the five coded values tied at 2 received the average rank of 10. The three coded values tied at 3 received the average rank of 14, and the six coded values tied at 4 received the average rank of 18.5.

Reconstitute the original sample sets of coded data of plants A, B, and C with the final tied ranks. In some applications there may be both individual ranks and tied ranks. Wherever there are tied ranks, they are to be used. Now do the following analysis for plant columns A, B, and C.

Kruskal-Wallis (Continued)

Example 8.31 (continued):

	Plant A	Plant B	Plant C
	4	4	4
	4	4	10
	4	4	14
	10	10	18.5
	10	14	18.5
	10	18.5	18.5
	14		18.5
	18.5		
Rank Sum	74.5	54.5	102.0
n	8	6	7
$(\text{Rank Sum})^2/n$	693.781	495.042	1486.286

$G = \Sigma(\text{Rank Sum})^2/n = 693.781 + 495.042 + 1486.286 = 2675.109$ $N = 8 + 6 + 7 = 21$

The significance statistic is H. H is distributed as chi square. Tie values are included in the calculation of chi square.

Let t = number of tied values in each tied set. Then $T = t^3 - t$ for that set.

Tied Set	t	T
1	7	336
2	5	120
3	3	24
4	6	210

Let $J = \Sigma T = 690$

$$H = \left(\frac{\frac{12}{N(N+1)} G - 3(N+1)}{1 - \frac{J}{(N^3 - N)}} \right) = \left(\frac{\frac{12}{(21)(22)}(2675.109) - 3(21+1)}{1 - \frac{690}{(21^3 - 21)}} \right) = 3.76$$

Let k = number of sample sets. $DF = k - 1 = 3 - 1 = 2$. Let $\alpha = 0.05$.

Critical chi square = $\chi^2_{0.05,\,2} = 5.99$

H is less than critical chi square. Therefore, the null hypothesis of equality of population medians cannot be rejected.

Mann-Whitney U Test

When there are ordinal measurements, the Mann-Whitney U test may be used to test whether two independent groups have been drawn from the same population. This is a powerful nonparametric test, and is an alternative to the t test when the normality of the population is either unknown, or believed to be non-normal.

Consider two populations, A and B. The null hypothesis, H_0, is that A and B have the same frequency distribution with the same shape and spread (the same median). An alternative hypothesis, H_1, is that A is larger than B, a directional hypothesis. We may accept H_1 if the probability is greater than 0.5 that a score from A is larger than a score from B. That is, if *a* is one observation from population A, and *b* is one observation from population B, then H_1 is that $P(a > b) > 0.5$.

If the evidence from the data supports H_1, this implies that the bulk of population A is higher than the bulk of population B. If we wished instead to test if B is statistically larger than A, then H_1 is $P(a > b) < 0.5$. For a 2-tailed test, that is, for a prediction of differences which does not state direction, H_1 would be $P(a > b) \neq 0.5$ (the medians are not the same).

If there are n_1 observations from population A, and n_2 observations from population B, rank all $(n_1 + n_2)$ observations in ascending order. Ties receive the average of their rank number. The data sets should be selected so that $n_1 \leq n_2$. Calculate the sum of observation ranks for population A, and designate the total as R_a, and the sum of observation ranks for population B, and designate the total as R_b.

$$U_a = n_1 n_2 + 0.5\, n_1 (n_1 + 1) - R_a$$
$$U_b = n_1 n_2 + 0.5\, n_2 (n_2 + 1) - R_b$$
$$\text{Where } U_a + U_b = n_1 n_2$$

Calculate the U statistic as the smaller of U_a and U_b. For $n_2 \leq 20$, Mann-Whitney tables are used to determine the probability, based on the U, n_1, and n_2 values. The tables up to $n_2 = 7$ are shown in Tables 8.38 and 8.39. This probability is then used to reject or fail to reject the null hypothesis. If $n_2 > 20$, the distribution of U rapidly approaches the normal distribution and the following apply:

$$U \text{ mean} = \mu_U = 0.5\, n_1 n_2$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \qquad Z = \frac{U - \mu_U}{\sigma_u} = \frac{U - 0.5\, n_1 n_2}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Mann-Whitney U Test (Continued)

Example 8.32: Consider an experimental group (E) and a control group (C) with scores as shown in Table 8.36. Note that $n_1 = 3$ and $n_2 = 4$. Does the experimental group have higher scores than the control group? H_0: A and B have the same median. H_1: median A is larger than median B. Accept H_1: if P (a > b) > 0.5.

E scores	10	12	17	
C scores	7	9	11	15

Table 8.36

To find U, we first rank the combined scores in ascending order, being careful to retain each score's identity as either an E or C, as in Table 8.37.

Rank	1	2	3	4	5	6	7	$R_{E,C}$	$U_{E,C}$
E			10		12		17	15	3
C	7	9		11		15		13	9

Table 8.37

The U_E and U_C were determined using the equations given on the prior page. U = minimum(U_E, U_C) = minimum(3, 9) = 3. The H_0 probability for $n_1 = 3$, $n_2 = 4$, and U = 3 is shown in Table 8.39 as P = 0.200. Since this is less than 0.5, we fail to reject H_0, and conclude that scores for both groups have come from the same population.

The probabilities in Table 8.38 are one-tailed. For a two-tailed test, the values for P shown in the Table should be doubled.

$n_2 = 3$			
	n_1		
U	1	2	3
0	0.250	0.100	0.050
1	0.500	0.200	0.100
2	0.750	0.400	0.200
3		0.600	0.350
4			0.500
5			0.650

Table 8.38 $n_2 = 3$

Note that this and the following Mann-Whitney Tables come from Siegel (1956)[22], which references the work of H. B. Mann and D. R. Whitney.

Mann-Whitney U Test (Continued)

$n_2 = 4$				
	n_1			
U	1	2	3	4
0	0.200	0.067	0.028	0.014
1	0.400	0.133	0.057	0.029
2	0.600	0.267	0.114	0.057
3		0.400	0.200	0.100
4		0.600	0.314	0.171
5			0.429	0.243
6			0.571	0.343
7				0.443
8				0.557

$n_2 = 5$					
	n_1				
U	1	2	3	4	5
0	0.167	0.047	0.018	0.008	0.004
1	0.333	0.095	0.036	0.016	0.008
2	0.500	0.190	0.071	0.032	0.016
3	0.667	0.286	0.125	0.056	0.028
4		0.429	0.196	0.095	0.048
5		0.571	0.286	0.143	0.075
6			0.393	0.206	0.111
7			0.500	0.278	0.155
8			0.607	0.365	0.210
9				0.452	0.274
10				0.548	0.345
11					0.421
12					0.500

$n_2 = 6$						
	n_1					
U	1	2	3	4	5	6
0	0.143	0.036	0.012	0.005	0.002	0.001
1	0.286	0.071	0.024	0.010	0.004	0.002
2	0.428	0.143	0.048	0.019	0.009	0.004
3	0.571	0.214	0.083	0.033	0.015	0.008
4		0.321	0.131	0.057	0.026	0.013
5		0.429	0.190	0.086	0.041	0.021
6		0.571	0.274	0.129	0.063	0.032
7			0.357	0.176	0.089	0.047
8			0.452	0.238	0.123	0.066
9			0.548	0.305	0.165	0.090
10				0.381	0.214	0.120
11				0.457	0.268	0.155
12				0.545	0.331	0.197
13					0.396	0.242
14					0.465	0.294
15					0.535	0.350
16						0.409
17						0.469
18						0.531

$n_2 = 7$							
	n_1						
U	1	2	3	4	5	6	7
0	0.125	0.028	0.008	0.003	0.001	0.001	0.000
1	0.250	0.056	0.017	0.006	0.003	0.001	0.001
2	0.375	0.111	0.033	0.012	0.005	0.002	0.001
3	0.500	0.167	0.058	0.021	0.009	0.004	0.002
4	0.625	0.250	0.092	0.036	0.015	0.007	0.003
5		0.333	0.133	0.055	0.024	0.011	0.006
6		0.444	0.192	0.082	0.037	0.017	0.009
7		0.556	0.258	0.115	0.053	0.026	0.013
8			0.333	0.158	0.074	0.037	0.019
9			0.417	0.206	0.101	0.051	0.027
10			0.500	0.264	0.134	0.069	0.036
11			0.583	0.324	0.172	0.090	0.049
12				0.394	0.216	0.117	0.064
13				0.464	0.265	0.147	0.082
14				0.538	0.319	0.183	0.104
15					0.378	0.223	0.130
16					0.438	0.267	0.159
17					0.500	0.314	0.191
18					0.562	0.365	0.228
19						0.418	0.267
20						0.473	0.310
21						0.527	0.355
22							0.402
23							0.451
24							0.500

Table 8.39 Additional Mann-Whitney Tables

Wilcoxon-Mann-Whitney Rank Sum Test

The Wilcoxon-Mann-Whitney rank sum test is similar in application to the Mann-Whitney Test. The null hypothesis is that the two independent random samples are from the same distribution. The alternate hypothesis is the two distributions are different is some way. Note that this test does not require normal distributions.

The observations or scores of the two samples (A and B) are combined in order of increasing rank and given a rank number. Tied values are assigned tied rank values. In cases where equal results occur, the mean of the available rank numbers is assigned. Next find the rank sum, R, of the smaller sample. Let N equal the size of the combined samples ($N = n_1 + n_2$) and n equal the size of the smaller sample. Then calculate:

$$R' = n\,(N + 1) - R$$

The rank sum values, R and R', are compared with critical values from Table 8.41 (Kanji, 1999)[15]. It represents critical values of the smaller rank sum. If either R or R' is less than the critical value, the null hypothesis of equal means is rejected. If $n_2 > 20$, the equations from the U test given on page VIII-85 are used for the Z calculation.

Example 8.33: Determine if the data from samples A and B (see Table 8.40) have the same distribution. The null hypothesis, H_0, is the data from samples A and B have the same median. The alternate hypothesis, H_1, is A median is larger than B median.

a	27	32	38	40	40	41	44	46	46		R
Rank	1	2	6	7	8	9	13	15.5	15.5		77
b	34	34	35	41	42	44	45	47	49	51	
Rank	3	4	5	10	11	12	14	17	18	19	113

Table 8.40

$n_A = 9$, $n_B = 10$, $N = 19$, $R = 77$ and
$$R' = n\,(N + 1) - R = (9)(20) - 77 = 103$$

Let $\alpha = 0.05$ for a one-tailed test. From Table 8.41 the critical value is 69. Since R = 77 is larger than 69, we fail to reject the null hypothesis of equal means. If H_1 had been A median is different than B median, then a two-tailed test would have been used.

Wilcoxon-Mann-Whitney Critical Values

Two-tailed		0.20	0.10	0.05	0.01
One-tailed		0.10	0.05	0.025	0.005
n_2	n_1				
3	2	3			
3	3	7	6		
4	2	3			
4	3	7	6		
4	4	13	11	10	
5	2	4	3		
5	3	8	7	6	
5	4	14	12	11	
5	5	20	19	17	15
6	2	4	3		
6	3	9	8	7	
6	4	15	13	12	10
6	5	22	20	18	16
6	6	30	28	26	13
7	2	4	3		
7	3	10	8	7	
7	4	16	14	13	10
7	5	23	21	20	16
7	6	32	29	27	24
7	7	41	39	36	32
8	2	5	4	3	
8	3	11	9	8	
8	4	17	15	14	11
8	5	25	23	21	17
8	6	34	31	29	25
8	7	44	41	38	34
8	8	55	51	49	43
9	1	1			
9	2	5	4	3	
9	3	11	9	8	6
9	4	19	16	14	11
9	5	27	24	22	18
9	6	36	33	31	26
9	7	46	43	40	35
9	8	58	54	51	45
9	9	70	66	62	56
10	1	1			
10	2	6	4	3	
10	3	12	10	9	6
10	4	20	17	15	12
10	5	28	26	23	19
10	6	38	35	32	27
10	7	49	45	42	37

Two-tailed		0.20	0.10	0.05	0.01
One-tailed		0.10	0.05	0.025	0.005
n2	n1				
10	8	60	56	53	47
10	9	73	69	65	58
10	10	87	82	78	71
11	1	1			
11	2	6	4	3	
11	3	13	11	9	6
11	4	21	18	16	12
11	5	30	27	24	20
11	6	40	37	34	28
11	7	51	47	44	38
11	8	63	59	55	49
11	9	76	72	68	61
11	10	91	86	81	73
11	11	106	100	96	87
12	1	1			
12	2	7	5	4	
12	3	14	11	10	7
12	4	22	19	17	13
12	5	32	28	26	21
12	6	42	38	35	30
12	7	54	49	46	40
12	8	66	62	58	51
12	9	80	75	71	63
12	10	94	89	84	76
12	11	110	104	99	90
12	12	127	120	115	105
13	1				
13	2	7	5	4	
13	3	15	12	10	7
13	4	23	20	18	14
13	5	33	30	27	22
13	6	44	40	37	31
13	7	56	52	48	44
13	8	69	64	60	53
13	9	83	78	73	65
13	10	98	92	88	79
13	11	114	108	103	93
13	12	131	125	119	109
13	13	149	142	136	125
14	1	1			
14	2	7	5	4	
14	3	16	13	11	7
14	4	25	21	19	14

Two-tailed		0.20	0.10	0.05	0.01
One-tailed		0.10	0.05	0.025	0.005
n2	n1				
14	5	35	31	28	22
14	6	46	42	38	32
14	7	59	54	50	43
14	8	72	67	62	54
14	9	86	81	76	67
14	10	102	96	91	81
14	11	118	112	106	96
14	12	136	129	123	112
14	13	154	147	141	129
14	14	174	166	160	147
15	1	1			
15	2	8	6	4	
15	3	16	13	11	8
15	4	26	22	20	15
15	5	37	33	29	23
15	6	48	44	40	33
15	7	61	56	52	44
15	8	75	69	65	56
15	9	90	84	79	69
15	10	106	99	94	84
15	11	123	116	110	99
15	12	141	133	127	115
15	13	159	152	145	133
15	14	179	171	164	151
15	15	200	192	184	171
16	1	1			
16	2	8	6	4	
16	3	17	14	12	8
16	4	27	24	21	15
16	5	38	34	30	24
16	6	50	46	42	34
16	7	64	58	54	46
16	8	78	72	67	58
16	9	93	87	82	72
16	10	109	103	97	86
16	11	127	120	113	102
16	12	145	138	131	119
16	13	165	156	150	130
16	14	185	176	169	155
16	15	206	197	190	175
16	16	229	219	211	196

Table 8.41 Wilcoxon-Mann-Whitney Critical Values

Nonparametric Test Summary

For tests of population location, the following nonparametric tests are analogous to the parametric t tests and analysis of variance procedures in that they are used to perform tests about population location or center value. The center value is the mean for parametric tests and the median for nonparametric tests.

- One-sample sign performs a test of the median and calculates the corresponding point estimate and confidence interval. Use this test as a nonparametric alternative to the one-sample Z and one-sample t tests.

- One-sample Wilcoxon performs a signed rank test of the median and calculates the corresponding point estimate and confidence interval. Use this test as a nonparametric alternative to the one-sample Z and one-sample t tests.

- Mann-Whitney performs a hypothesis test of the equality of two population medians and calculates the corresponding point estimate and confidence interval. Use this test as a nonparametric alternative to the two-sample t test.

- Kruskal-Wallis performs a hypothesis test of the equality of population medians for a one-way design (two or more populations). This test is a generalization of the procedure used by the Mann-Whitney test and, like Mood's median test, offers a nonparametric alternative to the one-way analysis of variance. The Kruskal-Wallis test looks for differences among the population medians.

- Mood's median test performs a hypothesis test of the equality of population medians in a one-way design. Mood's median test, like the Kruskal-Wallis test, provides a nonparametric alternative to the usual one-way analysis of variance. Mood's median test is sometimes called a median test or sign scores test.

The Kruskal-Wallis test is more powerful (the confidence interval is narrower, on average) than Mood's median test for analyzing data from many distributions, including data from the normal distribution, but is less robust against outliers.

Listed on the following page is a recap of some common nonparametric tests. It should be noted that nonparametric tests are less powerful (they require more data to find the same size difference) than the equivalent t tests or ANOVA tests. In general, nonparametric procedures are used either when parametric assumptions cannot be met, or when the nature of the data requires a nonparametric test.

Nonparametric Test Summary (Continued)

Test Name	Data Type	Test	Application
Kruskal-Wallis Test	Measurement or Count	X^2	Data is ranked or converted to ranks for a 1-way analysis of variance.
Kendall Coefficient of Concordance	Ranked Data	X^2	Determines degree of association among classifications based on ranked scores.
Spearman Rank Correlation Coefficient	Ranked Data	Formula	A measure of association (r_s) which requires both variables be measured on at least an ordinal scale.
Kendall Rank Correlation Coefficient	Ranked Data	Formula X^2	Same as Spearman, except it calculates r, which also permits determining partial correlation coefficient.
Contingency Coefficient	Count Data	X^2	A measure of association between two classifications.
Mann-Whitney U Test	Ranked Data	Tables	Determines if two independent groups are from same population. An alternative if t test assumptions cannot be met.
Wilcoxon-Mann-Whitney Rank Sum Test	Ranked Data	Tables	Same as above. Slightly simpler to use. An alternative if Z and t test assumptions cannot be met. It is equivalent to the Mann-Whitney U test.
Levene's Test	Converts Data to Squares of Deviations	t	Verifies homogeneity of variances across k samples. An alternative procedure if F and t assumptions are not met.
Mood's Median Test	Sample Medians	X^2	Determines equality of sample medians by scoring sample medians relative to population median.
Kolmogorov-Smirnov 1-sample	Sample Values	Table	Goodness-of-fit between observed scores and specified theoretical distribution.
Kolmogorov-Smirnov 2-sample	Sample Values	Table	Determines whether two independent samples come from the same distribution.
McNemar Test	Classification or Ranks	X^2	Determines significance of change in before and after designs.
Walsh Test	Response Scores	α% of 2^N	Randomization test for matched pairs.
Fisher Exact Probability	Classification or Ranks	Table or alpha	Determines whether two groups differ in the proportions within two classifications.
Cochran Q Test	Classification or Ranks	X^2	Determines whether 3 or more matched set of proportions differ significantly.
Friedman Test	Ranks	X^2	2-way analysis of variance for k matched samples.
Runs Test	Symbol Sequence	Table	Determines whether a sequence of samples are randomly distributed.

Table 8.42 Comparison Summary of Nonparametric Tests

Failure Mode and Effects Analysis

A FMEA provides the design engineer, reliability engineer, and others a systematic technique to analyze a system, subsystem, or item for all potential or possible failure modes. This method then places a probability that the failure mode will actually occur and what the effect of this failure is on the rest of the system. If criticality of failure is considered, the technique is called FMECA. The criticality portion of this method allows one to place a value or rating on the criticality of the failure effect on the entire system. It is not uncommon to omit the criticality portion from the methodology.

One of the controlling documents on this method is MIL-STD-1629, *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*[17]. In this standard, one will find modeling methods, definitions of severity classifications, and criticality values. One will also find sample forms and examples for performing a FMEA.

A FMEA or FMECA (in most cases there is little difference) is a detailed analysis of a system down to the component level. Once all items are classified as to the 1) failure mode, 2) effect of the failure, and 3) probability that failure will occur, they are rated as to their severity via an index called a RPN (risk priority number).

This RPN is dimensionless from the aspect that there is no real meaning to a value of say 600 versus 450 except in the difference in magnitude. Once all components or items have been analyzed and assigned a RPN value, it is common to work from the highest RPN value down. There will be more discussion on the RPN later.

FMEA Process Steps

The following are the steps that are taken to prepare a FMEA. An example of a completed FMEA is shown in Table 8.43.

1. FMEA number: This should be a log controlled number assigned by the reliability group for tracking the document.

2. The part number, name, or other appropriate description.

3. The design responsibility: Which department or group is responsible for this design?

4. The person responsible for FMEA preparation.

5. The date the FMEA was prepared and any necessary revision level.

FMEA Process Steps (Continued)

6. The subsystem or component part number being analyzed.

7. The component function.

8. The potential failure mode.

9. The potential effect of failure.

10. The potential cause of failure.

11. What are the current controls in place to prevent the cause from occurring?

Risk Assessment and RPN

The next major step is to weigh the risks associated with the current component, effect and cause, with the controls that are currently in place.

12. P is the probability this failure mode will occur. Values for this index generally index from 1 to 10 with 1 being virtually no chance and 10 being near certainty of occurrence.

13. S is the severity of the effect of the failure on the rest of the system if the failure occurs. These values are often indexed from 1 to 10. A value of 1 means the user will be unlikely to notice with a 10 meaning that the safety of the user is in jeopardy.

14. D is a measure of the effectiveness of the current controls (in place) to identify the potential weakness or failure prior to release to production. This index may also range from 1 to 10. A value of 1 means this will certainly be caught whereas a value of 10 indicates the design weakness would most certainly make it to final production without detection.

15. RPN. The risk priority number is the product of the indices from the previous three columns.

$$RPN = (P)(S)(D)$$

Risk Assessment and RPN (Continued)

16. **Actions are then based upon what items either have the highest RPN and/or have the major safety issues.**

17. **There is a column for actions to be taken to reduce the risk, a column for this responsibility and finally a column for the revised RPN once corrective action is implemented.**

In summary, FMEA provides a disciplined approach for the engineering team to evaluate designs to ensure that all the possible failure modes have been taken into consideration.

System FMECA

Part No./Name:	37XT11-Lock Mech.	P = Probability	FMEA No. 43		
Project:	Re-design	S = Seriousness	Final Design Deadline:	July 15, 2012	
Other Departments:	Shop Service, etc.	D = Likelihood	Prepared By:	RCD	
Subsystem Name:	Quill Clamping Mechanisms	RPN = Risk Priority Number	Reviewed By:	BLW	
Suppliers Involved:	Wilton and others		FMEA Date: 6-15-12	Rev.	
Design Responsibility:	Bob Dovich				

PART NUMBER NAME	FUNCTION	POTENTIAL FAILURE MODE(S)	POTENTIAL EFFECT(S) OF FAILURE	POTENTIAL CAUSE(S) OF FAILURE	CURRENT CONTROLS	RISK ASSESSMENT				RECOMMENDED CORRECTIVE ACTION(S)	ACTION(S) TAKEN	REVISED RISK ASSESSMENT				RESPONSIBLE DEPT OR INDIVIDUAL
						P	S	D	RPN			P	S	D	RPN	
WILTON POWER LOCK	CLAMP	LEAK	HOUSE-KEEPING	WEAR	ACCEPT SUPPLIER'S INFO	2	4	3	24	DISCUSS WITH SUPPLIER						
		LOSES CLAMPING FORCE (SHIFTING)	MACHINING PARTS OVERSIZE	SELECTED INADEQUATE SIZE POWER LOCK	ENG. STANDARD	2	4	4	32	PERFORM LOAD TESTS						
				MATERIALS & WORKMAN-SHIP	STD. Q.C.	1	4	2	8	NONE						
				OVER PRESSURE	NONE	2	4	2	16	REVIEW NEED FOR SYSTEM TO PREVENT OVER-PRESSURIZATION						
				PUMP SIZING	ENG. STANDARD	1	4	2	8	REVIEW PRESSURE DELIVERED IN FIELD AND ACTUAL NEED						

Table 8.43 An Illustrative FMEA

Risk Assessment

Risk assessment is the combination of the probability of an event or failure and the consequence(s) of that event or failure to a system's operators, users, or its environment. The analysis of risk of failure normally utilizes two measures of failure. These measures are:

- Severity of failure
  The effect of the failure on the system, operators, or mission

- Probability of failure *
  The likelihood of the failure occurring

The severity of failure is generally defined by the hazard severity categories from MIL-STD-1629, *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*[17]. These are shown in Table 8.44.

	Classification	Description
I	Catastrophic	A failure that may cause death or mission loss.
II	Critical	A failure that may cause severe injury or major system damage.
III	Marginal	A failure that may cause minor injury or degradation in mission performance.
IV	Minor	A failure that does not cause injury or system damage but may result in system failure and unscheduled maintenance.

Table 8.44 Hazard Severity Categories

The above categories were built with military missions in mind. Therefore, commercial companies should modify the definitions. The concept is that the potential result of the failure defines the hazard category.

* Note that the illustrative FMEA on the prior page used two probability components: The probability that the failure mode will occur (P) and the probability of detection prior to release (D).

Risk Assessment (Continued)

Another example, using the concept for commercial applications, is a severity index based on a scale of 1 to 10. (Ireson & Coombs)[12]

Ranking	Criteria
1	Unreasonable to expect that the minor nature of this failure will degrade the performance of the system.
2 - 3	Minor nature of failure will cause slight annoyance to the customer. The customer may notice a slight deterioration of the system performance.
4 - 6	Moderate failure will cause customer dissatisfaction. Customer will notice some system performance deterioration.
7 - 8	High degree of customer dissatisfaction and inoperability of the system. Does not involve safety or noncompliance to government regulations.
9 - 10	Very high severity ranking in terms of safety-related failures and nonconformance to regulations and standards.

Table 8.45 Commercial Severity Index (Scale 1 to 10)

The hazard classification or severity index is generated for each component or subsystem by the responsible party. This classification is based on the expected results of the failure of the component or subsystem. The probability of failure may also be ranked. A common ranking of failure probabilities is shown is Table 8.46.

Failure Probability Level	Description	Probability
A	High likelihood of occurrence	$>10^{-1}$
B	Probable occurrence	10^{-1} to 10^{-2}
C	Occasionally occurs	10^{-2} to 10^{-3}
D	Remote probability	10^{-3} to 10^{-6}
E	Highly unlikely	$<10^{-6}$

Table 8.46 A Common Failure Probability Ranking

Risk Assessment (Continued)

A number of systems are used to combine the probability of failure and the hazard category. These systems are based on accepting a degree of risk of occurrence with respect to the severity of the hazard. For instance, Table 8.47 shows one type of risk assessment matrix.

Hazard Category	Allowable Failure Probability Level
I Catastrophic	E (Unlikely)
II Critical	E (Unlikely)
III Marginal	D (Remote)
IV Minor	C (Occasional)

Table 8.47 Risk Assessment Matrix

Frequently, catastrophic failure modes have additional safety measures, such as redundant components, or frequent inspection during service, etc.

Note that no hazard category has an allowable failure probability of "frequent" or "probable."

Failure Mechanisms Versus Modes

The failure mode is the actual symptom of the failure. That is, the failure mode may be premature engine shut-down, or 70% degradation of function, or any other description of what external occurrence will be defined as a failure. These failure modes are the result of failure mechanisms.

Failure mechanisms are the individual, or multiple reasons that cause the failure mode. For instance, failure mechanisms might be corrosion, or contamination, or any other description of reasons that might cause a failure mode. Failure mechanisms cause failure modes.

Failure mech causes failure mode

Types of FMEAs

According to Bowles (1994)[5], there are four types of FMEAs:

- Design FMEAs
- Process FMEAs
- System FMEAs
- Functional FMEAs

The following is an abbreviated description of the types:

Design FMEAs are performed on the product or service (system) at the design level. The purpose is to analyze how failure modes affect the system and to minimize failure effects upon the system. Design FMEAs are used before products are released to the manufacturing operation. All anticipated design deficiencies should be detected and corrected by the end of this process.

Process FMEAs are performed on the manufacturing processes. They are conducted through the quality planning phase as an aid during production. The possible failure modes in the manufacturing process, limitations in equipment, tooling, gauges, operator training, or potential sources of error are highlighted, and corrective action is taken.

System FMEAs comprise part level FMEAs (RAC *Blueprints for Product Reliability - RBPR-3*, 1996)[4]. All of the part level FMEAs will tie together to form the system. As a FMEA (part level FMEA) goes lower into the system, into more detail, more failure modes will be considered. A system FMEA needs only to go down to the appropriate level of detail.

Functional FMEAs are also known as black box FMEAs. This type of FMEA focuses on the performance of the intended part or device, rather than on the specific characteristics of the individual parts. As an example, if a project is in the early design stages, a black box analysis would focus on the function of a device rather than on the exact specifications (color must be blue-gray, knob is 2.15 mm to the left, etc.)

All of the above FMEAs can be applied to software systems.

Additional Analysis Methods

Additional Analysis Methods are reviewed in the following topic areas:

- Gap analysis
- Root cause analysis
- Waste analysis

Gap Analysis

Gap analysis is an assessment tool to compare an organization's current performance to a desired or potential performance. The objective of a gap analysis is to identify the difference between what is and what should be. Refer to Figure 8.48 below:

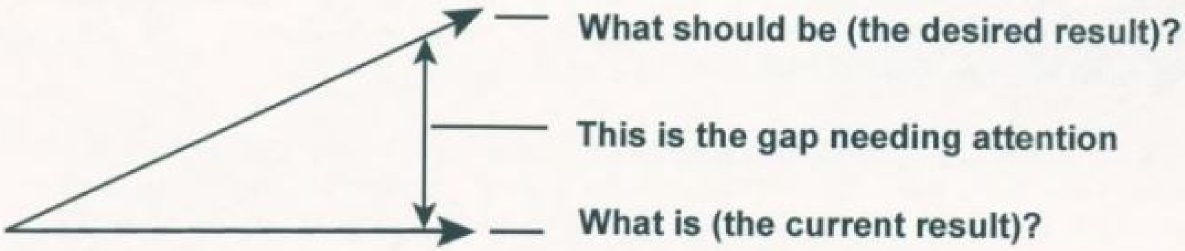


Figure 8.48 The Organizational Performance Gap

Gap analysis provides an organization with the mandate to direct resources to the appropriate areas and thereby achieve a desirable objective. Examples of such objectives are:

- Stay in business
- Retain or improve market share
- Improve employee moral
- Match or exceed the best external benchmarks
- Match or exceed the current competition
- Reduce cycle times
- Improve customer satisfaction
- Lower internal costs
- Achieve ISO 9001 certification
- Increase productivity
- Improve quality levels

Gap Analysis (Continued)

Any form of gap analysis requires a minimum of three categories of information:

- Where are we now?
- Where do we want to be?
- How are we going to measure the results?

Somewhat counterintuitively, this Primer will address the above items in a slightly different sequence by discussing the "where do we want to be" tools first.

Scenario Planning

When developing strategic plans, management may be prone to groupthink, overconfidence, tunnel vision, etc. In order to overcome potential errors in decision-making, best and worst case scenarios should be considered (collapse of the Berlin wall, rise of the internet, etc.). Once the different scenarios are visualized, appropriate contingency plans can be developed.

Scenario planning is a strategic planning process in which a relatively small number of futuristic stories or scenarios are developed. These scenarios depict how an organization's future might unfold, and then proceeds to develop the best approaches for addressing these events. Organizational scenario planning is considered an adaptation from military intelligence in that action in warfare must often be initiated with incomplete or inadequate information.

All aspects of scenario planning are difficult to formalize because many essential elements such as environmental concerns, customer perceptions, resource depletions, and competition reactions are often unknown.

Scenario planning recognizes that there can certainly be subjectivity in the interpretation of facts or data. The scenario planning technique is a group process which encourages an open knowledge exchange surrounding the central issues important to an organization. The goal is to develop a number of potential responses to the perceived influential corporate risks. The method is widely used as a strategic management tool.

Scenario Planning (Continued)

The scenario planning process follows a sequence similar to that described below:

- Select people who will contribute a wide range of perspectives
- Develop a list of perceived social, technical, or economic shifts
- Cluster these views into related patterns (groupings)
- Develop a list of the best views (priorities)
- Develop a rough picture of the future based on these priorities
- Determine how each scenario will affect the organization
- Determine potential courses of action to take
- Identify early warning signs indicating scenario arrival
- Monitor, evaluate, and review the scenarios

Often a list of 6 - 10 perceived threats or opportunities are condensed into 2 or 3 developed scenarios (priorities).

There are certainly some traps to avoid in scenario planning:

- Not using an experienced facilitator
- Considering scenarios to be forecasts
- Making scenarios too simplistic
- Limiting the global impact on scenarios
- Failing to include an executive team in the process
- Treating scenarios as only an informational activity
- Limiting the imaginative stimulus into the scenario design
- Failing to develop scenarios for key business impact areas

(VBM, 2007)[25]

Hoshin Planning

Hoshin kanri or hoshin planning is an execution tool, that is used both to organize and deploy strategic plans throughout an organization. A disciplined management system is used for the strategic priorities. The Japanese word hoshin can be divided into two parts: ho means "method" and shin means "shiny metal showing direction." Kanri means "planning." Putting these phrases together indicates a methodology for setting a strategic direction, or in other words: "a management compass."

As a management tool, hoshin planning translates the company's vision into dramatic measurable results and strategic breakthroughs. The focus of hoshin planning is to identify the vital few breakthrough achievements. In a nutshell, hoshin planning has six basic objectives:

- Align the organizational goals
- Focus on the vital few strategic gaps
- Work with others to close the gaps
- Specify the methods and measures to achieve the strategic objectives
- Make the linkage among local plans visible
- Continuously improve the planning process (Bechtell, 1995)[2]

Although hoshin planning is a structured technique for planning and deployment, management must be committed to the techniques for effective results. It should be noted that the methods to achieve the results are important, not just the end results.

Other Key Analysis Techniques

The student should note that a number of other widely used techniques can also be employed for gap analysis. In this Primer, there is additional coverage of the following applicable topics:

- Benchmarking, reviewed in Section III
- SWOT analysis, reviewed in Section II
- PEST analysis, reviewed in Section II

The following pages contain discussions of organizational assessment and organizational metrics. These two items address the gap analysis questions of "Where are we now?" and "How are we going to measure the results?"

Organizational Assessments

An organizational assessment typically consists of a functional analysis, which includes data collection. Examples of data collection techniques are:

- Face-to-face interviews with management and workers
- Surveys of an appropriate organizational sample
- Focus group inputs
- Observations gathered through interviews and plant walk throughs
- Data gathered from industry sources

Structured assessment forms may vary from firm to firm. These assessment forms are available from governmental organizations and other sources. An assessment and analysis typically divides an organization into key functional areas. A summary of the important areas would include:

- Leadership: Determine the nature and style of organizational leadership. Are managers proactive or reactive, energetic or laid back, type X or type Y, etc.?

- Business practices: What is the organizational vision, mission, competitive forces, identified strengths, weaknesses, opportunities, and threats? What is the organizational structure and how wide is a manager's span of control?

- Financial analysis: Determine the company's financial ratios compared to industry averages. The ratios include: current assets to current liabilities, net sales to net worth, net sales to inventory, sales dollars per employee, fixed assets to net worth, etc. An analysis of the company's financial statements and cash flows may be included.

- Marketing: Identify customers and markets. Assess market plans and marketing efforts.

- Quality management: Does the organization practice defect detection or defect prevention measures? What are the quality systems (ISO 9001, ISO/TS 16949, etc.)? Does the company use cost of quality reports, scrap analysis, or SPC? Is the MBNQA, Shingo prize, or Deming prize important?

- Research & design: Determine the amount of concurrent engineering, number of new products, development cycle times, etc.

Organizational Assessments (Continued)

- Manufacturing: What is the state of automation, robotics, and material handling? What is the state of technology (leading edge or market follower)? What are the plant capabilities? What is the state of machinery repair or disrepair?

- Safety & health: Analyze safety committees, personnel training, hazard communications, respiratory equipment, lockout and tagout procedures, incident rates, active safety programs, housekeeping, inspections, etc.

- Information technology: Determine the uses of information technology (IT), electronic data interchange with customers and suppliers, electronic e-mail communications, data collection and analysis, PC programs, B2B usage, e-commerce, etc.

- Production engineering: Determine technical problem solving capabilities. Are engineering responses proactive or reactive?

- Production planning & scheduling: What are the material and inventory levels, scheduling capabilities, usage of scheduling software, delivery dates, actual ship dates, product tracking capabilities, etc.?

- Plant facilities: How many plant sites? Are facility conditions adequate? Is expansion required?

- Human resources: Analyze the needs of the work force. What are the training capabilities? What are employee turnovers, absent rates, and attitudes?

Obviously, variations in the above assessment areas are not only possible but probable. The detail of assessment items may vary considerably. The results of an assessment may be analyzed and presented to management. For effective use of the assessment, clear action plans must be developed and implemented. Unfortunately, many action plans will be dropped at this point and many assessment reports will end up on a bookshelf.

The assessor may or may not analyze the data or recommend solutions. Typically, he/she gathers the data and summarizes it into major categories for review by the management group. The aim is to have this group develop their own solutions. Thus, management will have ownership of the solutions, and be committed to implement the solutions and nurture them to success.

Organizational Metrics

Installing organizational metrics normally means that a set of performance goals and standards have been determined. Metric analysis should then provide effective control feedback for reaching these goals. Organizational performance goals and corresponding measurements are often established in the areas of:

- Profit
- Cycle times
- Resources
- Marketplace responses

(Besterfield, 1999)[3]

A company should develop metrics for each major performance goal. A unit and a method of measurement must be defined. For the above performance goals, possible metrics include:

Profit

- Stockholder value
- Capital investment
- Personnel costs
- Competitor comparisons
- Return on investment
- Sales dollars

Profit may be short-term (6 months or less) or long-term (2 years or more).

Cycle Times

- Existing cycle times
- Internal benchmarks
- External benchmarks
- Reductions in cycle times

Ten fold reductions in cycle times are possible.

Resources

- Number of improvement projects
- Return on improvement projects
- Process capability studies
- Variation reductions
- Costs of poor quality compared to some base
- Percent defects compared to some base

Organizational Metrics (Continued)

Marketplace Responses

- Market survey of customers
- Analysis of returns
- New product development
- Customer retention
- Customer losses
- Courtesy ratings
- Facilities ratings

Metrics can be and must be developed in order to measure achievement of the organizational goals. Dr. Deming describes the problem of obtaining a "true value." There is variation in all measurement and one must be skeptical of how data is collected.

The device used to collect the data must be accurate. Additionally, the questions of when, where, and how will impact the accuracy and precision of data. One should treat all data with caution. According to Juran (1993)[13], the development of any measurement system should take into account the following factors:

- There should be a standardized meaning of the measurement
- The data should help the decision-making process
- It should provide worthwhile information
- It should be easy to install
- If worthwhile, it should be benchmarked or used elsewhere

Any mechanical and electrical instruments, gages, tools, etc., used for data collection must undergo recognized calibration procedures. In many applications, the appropriate metrics are qualitative based on customer, supplier, or internal appraisal feedback forms.

Root Cause Analysis

An individual or team is often given the responsibility of determining the root cause of a deficiency and correcting it. The solution to some problems may be complex and difficult. In other cases, the solution may be known, but considerable time will be required to implement it. The proposed action may take several steps. Refer to Table 8.49 below:

Situation	Immediate Action	Intermediate Action	Root Cause Action
The dam leaks	Plug it	Patch the dam	Find out what caused the leak so it does not happen again. Then rebuild the dam.
Parts are oversized	100% inspection	Put an oversize kickout device in line	Analyze the process and take action to eliminate the production of oversized parts.

Table 8.49 Short and Long-Term Corrective Actions

To help locate the system's true problem, a variety of problem solving tools are available. Listed below are commonly used techniques.

Subjective Tools

Ask why, and then ask why again....
Brainstorming
Process flow analysis
Plan-do-check-act
Systematic problem solving
Nominal group technique
Operator observation
Fishbone diagrams
Consensus techniques
Six thinking hats
Use of teams
FMEA and FTA

Analytical Tools

Data collection and analysis
Pareto analysis
Regression analysis
Check sheets
Data matrix analysis
Process capability analysis
Partitioning of variation
Subgrouping of data
Simple trials
Statistical design of experiments
Analytical tests (X^2, F, Z, t)
Control charting

When permanent corrective action is proposed, management must determine if:

- The root cause analysis has identified the full extent of the problem
- The corrective action is satisfactory to eliminate or prevent recurrence
- The corrective action is realistic and maintainable

5 Whys

The 5 whys approach to root cause analysis is described as asking the question "Why?" five times. This technique is generally attributed to a Japanese method of determining the root cause. The following is an example of 5 whys.

Symptom: The customer shipments were not made on time.

1. Why? We ran out of parts because the die stamping press broke down.
2. Why? The press had not received scheduled maintenance for a period of three months.
3. Why? The maintenance department staff had been reduced to six people from eight.
4. Why? The maintenance department was over budget due to high overtime costs, and the General Manager eliminated all overtime and required a 25% reduction in personnel for all overhead support departments.
5. Why? The company was not reaching profit goals so the CEO had issued an order to avoid all unnecessary spending. So the root cause was the CEO was worried about getting fired for poor profit performance.

There is nothing magical about 5 whys. In fact, the root cause may be found after 3 or 4 whys. In other cases, one may need to ask "why?" six or more times. This method asks "why?" until the root cause is found.

5 Ws and H

The 5 Ws and H approach to root cause analysis is summarized as asking the questions Who?, What?, When?, Where?, Why?, and How? The 5 Ws and H methodology is an old technique, used by newspaper reporters in asking questions in order to get the full story. In the quality context, responses to these questions can be organized with each element depicted as a branch on a cause and effect (fishbone) diagram.

In some references, this same basic method is simply referred to as the 5 Ws (who, what, where, when, and why). Also note that the order of the Ws varies, depending upon the referenced source. The technique looks at a problem or symptom from several viewpoints in order to include as much information as is needed to assist in determining the root cause.

Cause and Effect (Fishbone) Diagrams

Cause and effect diagrams are effective team based tools to determine the potential root causes of a problem.

A cause and effect (fishbone) diagram:

- Breaks problems down into bite-size pieces
- Displays many possible causes in a graphic manner
- Is also called a cause and effect, 4-M or Ishikawa diagram
- Shows how various causes interact
- Follows brainstorming rules when generating ideas

A fishbone session is divided into three parts: brainstorming, prioritizing, and development of an action plan. Identify the problem statement and brainstorm the categories in a fishbone diagram. To prioritize problem causes, polling is often used. The three most probable causes may be circled for the development of an action plan.

Generally, the 4-M (manpower, material, method, machine) version of the fishbone diagram will suffice. Occasionally, the expanded version must be used. In a laboratory environment, measurement is a key issue. When discussing the brown grass in the lawn, environment is important. A 5-M and E schematic is shown in Figure 8.50.

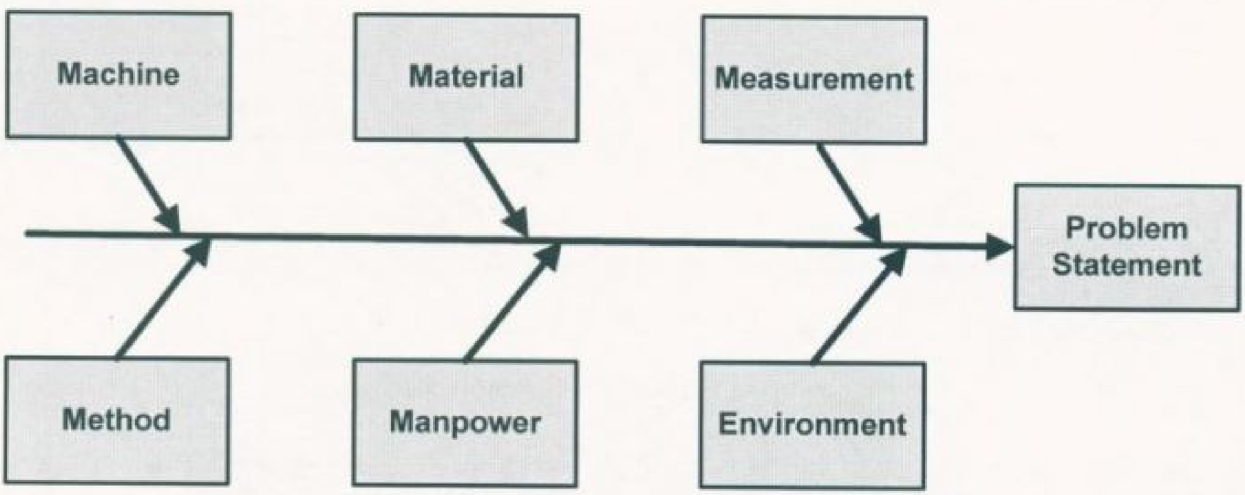


Figure 8.50 Basic Fishbone 5-M and E Example

Cause and Effect (Fishbone) Diagrams (Continued)

Shown below is an actual 5M and E schematic used by an improvement team to minimize part count error.

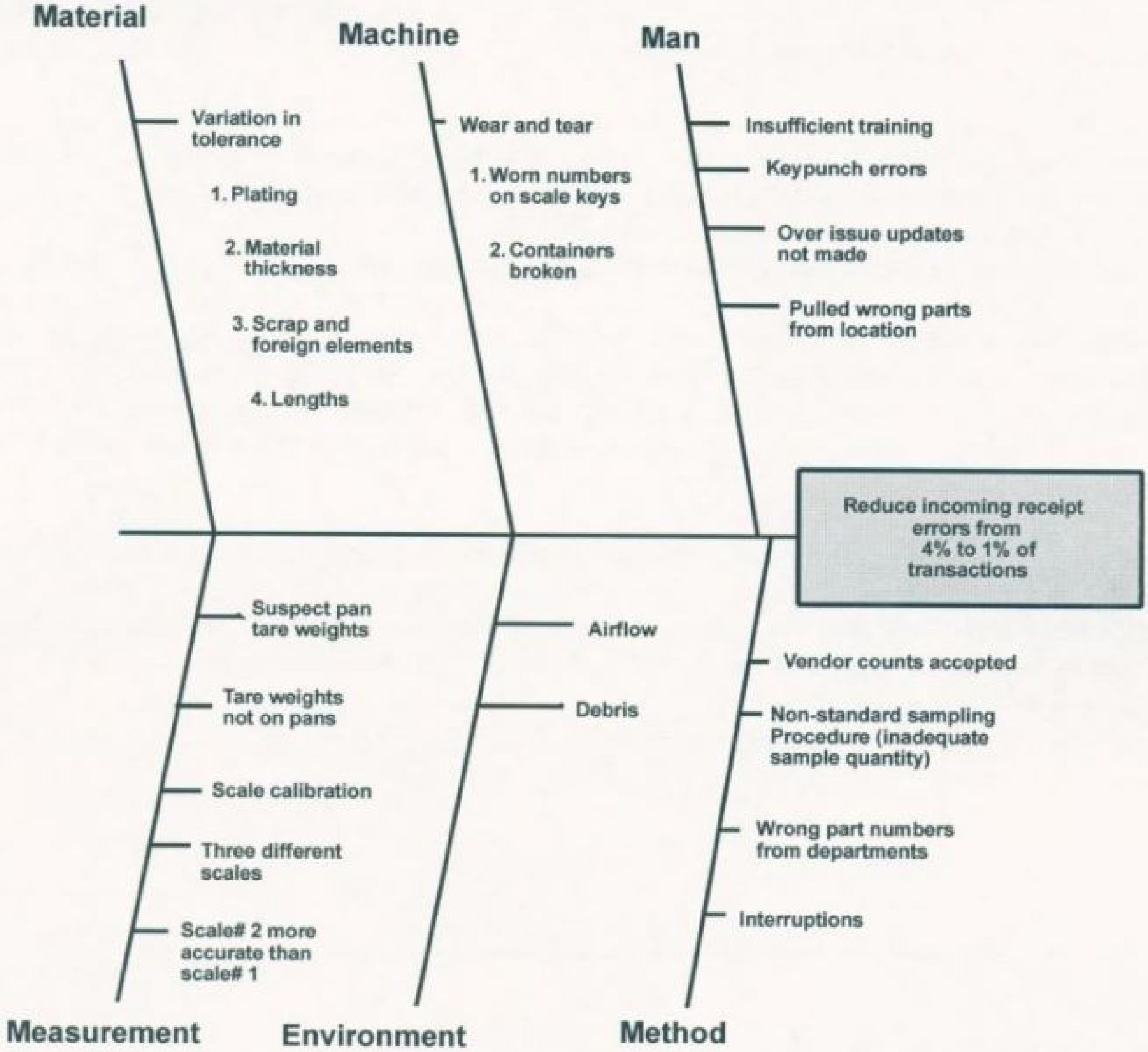


Figure 8.51 Actual Fishbone Example

For additional examples of cause and effect or Ishikawa diagrams refer to: Ishikawa, K., *Guide to Quality Control*, (1982)[11]. Dr. Ishikawa attributed the first application of a cause and effect diagram to Tomiko Hashimoto's article, "Elimination of Volume Rotation Defects Through QC Circle Activities," Factory Work and QC, No. 33. (Hashimoto)[8].

Pareto Diagrams

Pareto diagrams are very specialized forms of column graphs. They are used to prioritize problems (or opportunities) so that the major problems (or opportunities) can be identified. Pareto diagrams can help six sigma teams get a clear picture of where the greatest contribution can be made.

History

There is a very interesting story behind the name of Pareto diagrams. The word "Pareto" comes from Vilfredo Pareto (1848-1923). He was born in Paris after his family had fled from Genoa, Italy, in search of more political freedom. Pareto, an economist, made extensive studies about the unequal distribution of wealth and formulated mathematical models to quantify this maldistribution. He also wrote a political book on nationalism which helped lead to fascism in Italy.

Dr. Joe M. Juran, world renowned leader in the quality field, was preparing the *Quality Control Handbook*[14] in the late 1940s. He needed a short name to apply to the phenomenon of the "vital few" and the "trivial many." He depicted some cumulative curves in this manuscript and put a caption under them, "Pareto's principle of unequal distribution..." The text makes it clear that Pareto only applied this principle in his studies of income and wealth; Dr. Juran applied this principle as a "universal." Thus, the diagram could be named a "Juran diagram." To complicate matters more, the cumulative curve diagram itself was first used by M. O. Lorenz in 1904.

Vilfredo Pareto
1848-1923

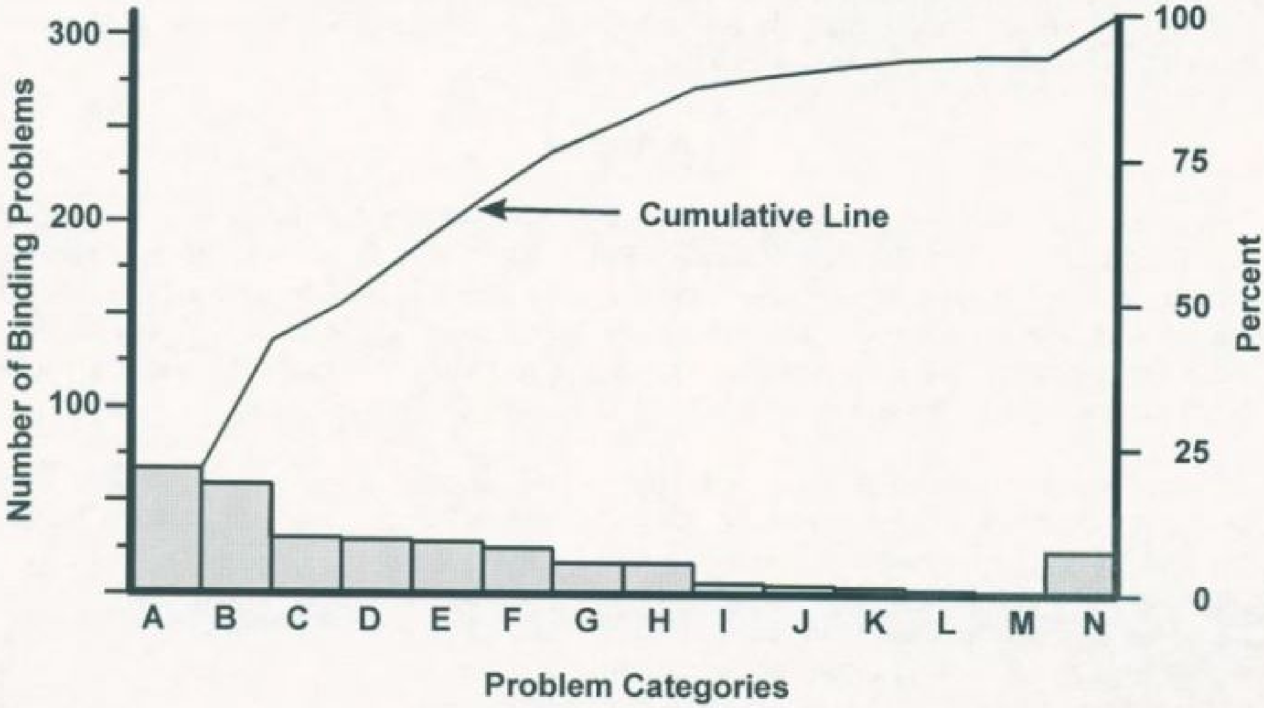
Briefly stated, the principle suggests (in most situations) that a few problem categories (approximately 20%) will present the most opportunity for improvement (approximately 80%).

Pareto diagrams are used to:

- Analyze a problem from a new perspective
- Focus attention on problems in priority order
- Compare data changes during different time periods
- Provide a basis for the construction of a cumulative line

Typical Pareto Diagram

The defects for a book product are shown in Pareto form below:



A. Emulsion - glue 67	H. Square variation 17
B. Grease/oil/dirt 59	I. Head bands 6
C. Hot melt - glue 30	J. Case damage - unknown 5
D. Sewing thread 29	K. Case damage - area II 4
E. Gilding defects 28	L. Upside down books 2
F. End sheet problems 25	M. Torn pages 0
G. Case damage - area I 17	N. All others 23

Figure 8.52 Typical Pareto Diagram

Note that the "all others" category is placed last. Cumulative lines are convenient for answering such questions as, "What defect classes constitute 70% of all defects?"

"First things first" is the thought behind the Pareto diagram. Our attention is focused on problems in priority order. The simple process of arranging data may suggest something of importance that would otherwise have gone unnoticed. Selecting classifications, tabulating data, ordering data, and constructing the Pareto diagram have proved to serve a useful purpose in problem investigation.

Weighted Pareto Analysis

Pareto analysis is very helpful in assisting improvement teams with the selection of serious problems. The Pareto method assumes that there will be segregation of the significant few from the trivial many. In many cases, the Pareto diagram is constructed based upon the number of event occurrences. However, criticality (or potential safety or economic loss) factors might result in a different Pareto alignment. Consider the following audit report results: (Anderson, 2012)[1]

	Deficiency Description	Occurrences
A	Critical test equipment overdue for calibration	1
B	Inadequate training of employees	1
C	Marginal internal auditing	1
D	Marginal procedures	12
E	Obsolete drawings in use in manufacturing	3
F	Problems with record retention	8
G	Unauthorized changes on controlled documents	4
H	Questionable criteria for inspection judgments	2
I	Inadequate corrective action	1
J	Unidentified material in stockroom	6
K	Employees not following procedures	2
L	Nonconforming material mixed with good	2

The traditional Pareto diagram, based on occurrences, would look like so:

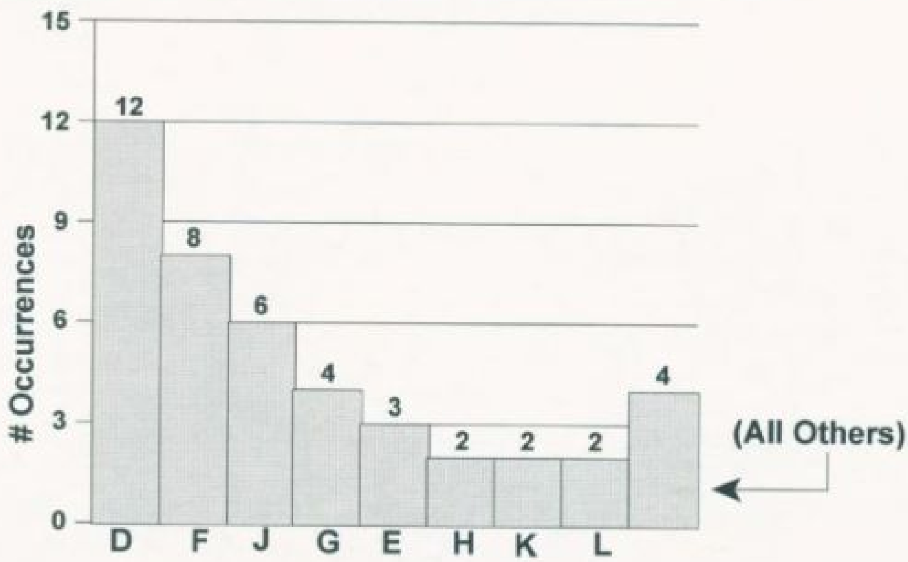


Figure 8.53 Audit Occurrence Pareto Diagram

Weighted Pareto Analysis (Continued)

Assume that an objective assessment of the criticality of the same deficiencies yielded the following weighting:

Critical (100 demerits each)	A, G, L
Major (25 demerits each)	C, E, K
Minor (10 demerits each)	B, D, I
Incidental (1 demerit each)	F, H, J

The accumulated, weighted audit data now becomes:

A	1 X 100 = 100	G	4 X 100 = 400
B	1 X 10 = 10	H	2 X 1 = 2
C	1 X 25 = 25	I	1 X 10 = 10
D	12 X 10 = 120	J	6 X 1 = 6
E	3 X 25 = 75	K	2 X 25 = 50
F	8 X 1 = 8	L	2 X 100 = 200

The composite Pareto diagram suggests a different priority than the previous one:

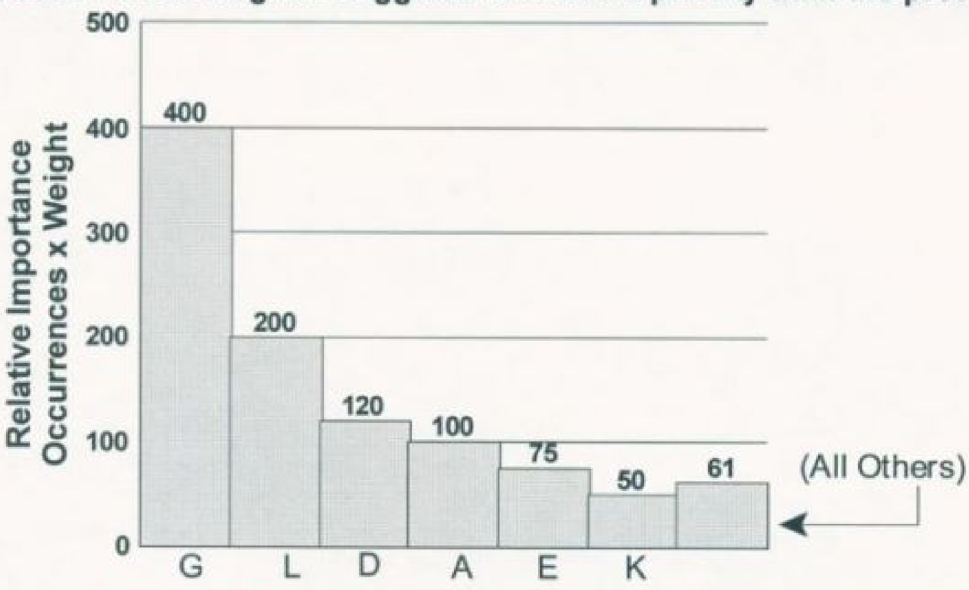


Figure 8.54 Criticality Weighted Pareto Diagram

Dollars can also be used instead of demerits. In addition, the original data could be sorted by shift, or by day of week, etc.

Ford 8D Problem Solving (Eight Disciplines)

The Ford Motor Company has developed a method of root cause analysis for use in group problem solving. It is a staged problem solving methodology. The project leader uses a checklist to ensure completion of each step. A disciplined approach should result in a satisfactory solution. The 8D process has the following steps:

D1. Establish the team: Use cross functional team membership. This is a diverse team with the knowledge, time, authority, and skill to solve the problem and implement the corrective action. There will be a team champion.

D2. Describe the problem: Identify the problem in terms of the internal/external customer problem. Define the problem in terms of "what is wrong with what?," use of the 5W 2H method (who, what, where, when, why, how, and how many). Use quantifiable terms to define the problem. Some other useful tools include: is/is not, cause and effect diagrams, and flow charts.

D3. Develop a containment action: Determine a short-term (interim) solution to the immediate problem. This prevents problems from affecting the customer.

D4. Identify the root cause: Search for all possible causes of the problem. Update is/is not, cause and effect diagram, process flow charts, etc. A FMEA (failure mode effect analysis) could also be used here. Test each potential root cause against the problem description and test data for elimination of the problem.

D5. Develop alternative solutions: Identify alternative corrective actions to eliminate the root cause. One must establish a link between the alternative solution and the root cause. The corrective action should not have any undesirable side affects. Analyze the risks involved.

D6. Implement a permanent corrective action: Provide a plan to install a permanent corrective action, with control monitoring to ensure elimination. Determine a schedule necessary to measure the problem. Evaluate and verify the effectiveness of the permanent corrective action.

D7. Prevent recurrence: Modify the existing systems, practices, and procedures to prevent recurrence of the problem. One must be able to state that this is the best possible long-term solution for the customer.

D8. Recognize team and individual contributions: Acknowledge team and individual contributions and celebrate.

Fault Tree Analysis

Fault tree analysis (FTA) is a systematic, deductive methodology for defining a single, specific, undesirable event, and determining all possible reasons (failures) that could cause that event to occur. The undesired event constitutes the top event in a fault tree diagram and generally represents a complete failure of the product. The FTA is an easier and faster method of analysis compared to FMEA because it focuses on a selected subset of all possible system failures, specifically those that can cause a catastrophic top event. FMEA progresses sequentially through all possible system failure modes, regardless of their severity.

When properly applied, a FTA is extremely useful during the initial product design phase as an evaluation tool for driving preliminary design modifications. Other potential uses of FTA include:

- Functional analysis of highly complex systems
- Evaluation of subsystem events on the top event
- Evaluation of safety requirements and specifications
- Evaluation of system reliability
- Identification of potential design defects and safety hazards
- Evaluation of potential corrective actions
- Maintenance and troubleshooting simplification
- Logical elimination of causes for an observed failure

FTA is preferred over FMEA when:

- The safety of public, operating, or maintenance personnel is paramount
- A small number of clearly differentiated top events can be identified
- Completion of a functional profile is of critical importance
- There is a high potential for failure due to human or software errors
- The primary concern is a quantified risk evaluation
- Product functionality is highly complex or highly interconnected
- The product is not repairable once initiated (space systems)

FMEA is preferred over FTA when:

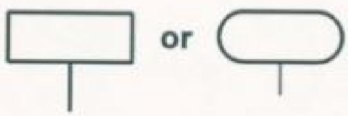
- Top events cannot be explicitly defined or limited to a small number
- Multiple potentially successful functional profiles are feasible
- The identification of all possible failure modes is important
- Product functionally has little human or software intervention

(Reliability Toolkit, 1993)[21]

FTA Symbols

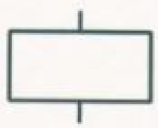
Fault tree analysis uses the concepts of logic gates to determine the overall reliability of a system. Fault tree analysis is also used in assessing potential system failure modes. There are numerous FTA symbols. These are broken down into two main categories; event symbols and gate symbols. Only a few examples are described below:

Event Symbols



Top event: Contains a description of a system-level fault or undesired event.

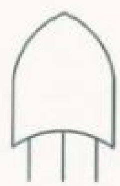
Basic event: Usually the lowest level of event fault that one wishes to study. It is used as an input to a logic gate.



Fault event: Contains a description of a lower-level fault. It can receive inputs from or provide outputs to a logic gate.

Logic (Gate) Symbols

"and" gate (series) — The output event occurs only if all the input events occur simultaneously.



"or" gate (parallel) — The output event occurs if any one of the input events occur.

Figure 8.55 Event and Logic Symbols and Descriptions

Fault Tree Analysis Example

Fault tree analysis begins at the system level, and assumes that a failure will occur. For instance, consider the hypothetical probabilities for a home computer failing to work as shown in Figure 8.56 below:

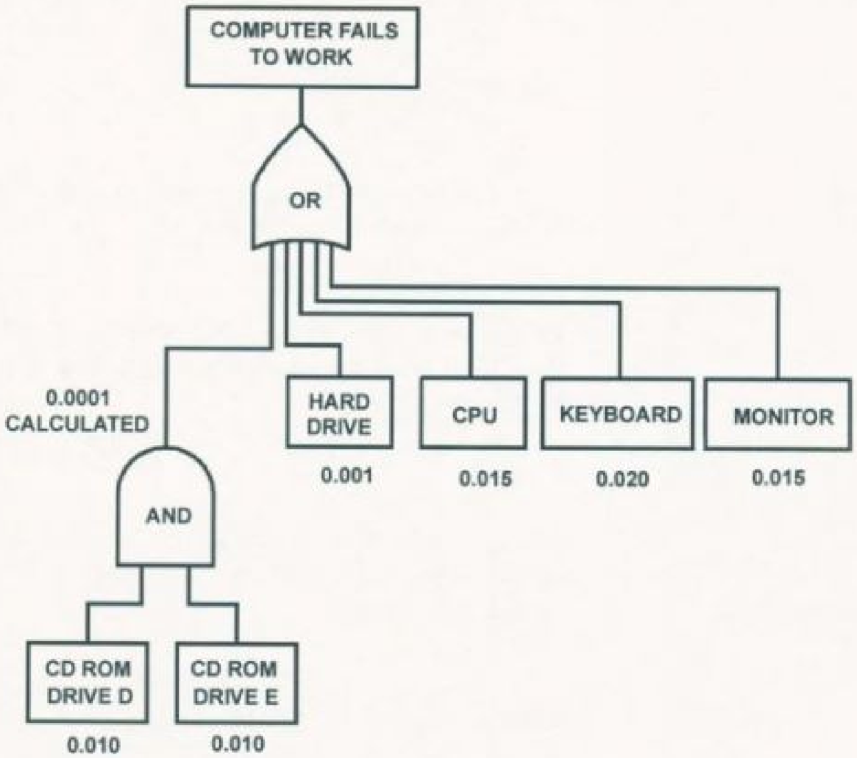


Figure 8.56 A Simple FTA Schematic

The fault tree in the above case indicates that the computer will fail to operate if CD ROM drive D <u>and</u> drive E fail, <u>or</u> the hard drive fails, <u>or</u> the CPU fails, <u>or</u> the keyboard fails, <u>or</u> the monitor fails. Since "and" gates multiply and "or" gates add, the probability of the home computer not working can be computed.

$$\mu_{system} = 1 - (1 - 0.0001)(1 - 0.001)(1 - 0.015)(1 - 0.020)(1 - 0.015)$$
$$\mu_{system} = 1 - (0.9999)(0.999)(0.985)(0.98)(0.985)$$
$$\mu_{system} = 1 - 0.9498 = 0.0502$$

The probability of a failure is 5.02%

As a root cause analysis tool, this fault tree indicates that the keyboard should receive priority attention (0.20) followed by the CPU (0.015) and monitor (0.015). Improvements in the hard drive and CD ROM drives can wait.

Waste Analysis

Non-value added activities are classified as muda. This term describes the waste that exists in a process. At each process step work is applied. The useful activities that the customer will pay for is considered value added. The other activities are not important to the customer or contain elements that the customer will not pay for. Those non-paying activities are muda. Imai (1997)[10] provides a list of seven waste categories that have been widely used:

- Overproduction
- Inventory
- Repair/rejects
- Motion
- Transport
- Processing
- Waiting

Overproduction

The muda of overproduction is producing too much at a particular point in time. Overproduction is characterized by:

- Producing more than is needed by the next process or customer
- Producing earlier than needed by the next process or customer
- Producing faster than is needed by the next process or customer

In the just-in-time environment, producing too early is as bad as producing too late. Parts need to be available at a certain location at a certain time according to the customer's schedule. Having the product too early, too late, or in quantities that are too great will result in undesirable consequences, such as:

- Extra space used at the customer's plant
- Extra space used at the organization's plant
- Extra raw materials in use
- Extra utilities used
- Extra transportation for the customer and organization
- Extra scheduling costs

Waste Analysis (Continued)

Inventory

Parts, raw materials, work-in-process (semi-finished goods), inventory, supplies, and finished goods are all forms of inventory. Excess inventory is considered muda since it does not add value to the product. Inventory will:

- Require space in the shop
- Require transportation
- Require forklifts
- Require conveyor systems
- Require additional labor
- Require interest on material costs

Inventory sitting around in various process stages can be adversely affected in the following ways:

- It may gather dust
- It can deteriorate
- It can become obsolete
- It may get wet
- It can experience handling damage

Repair/ Rejects

The repair or rework of defective parts involves a second attempt at producing a good item. Scrapping the whole part is a definite waste of resources. Having rejects on a continuous flow line defeats the purpose of continuous flow. Line operators and maintenance will be used to correct problems, putting the takt time off course. Repair and rework may require nonconforming product forms to be filled out by suppliers, generating additional muda.

Various design changes can be muda also. Design changes are often considered rework or an extra development effort. Both of these activities will create the need for additional labor.

Waste Analysis (Continued)

Motion

The efficient use of the human body is critical to the well being of the operator. Extra unneeded motions are wasteful. Operators should not have to walk excessively, lift heavy loads, bend awkwardly, reach too far, repeat motions, etc. New tools should be designed to help with strenuous hand or body motions. The layout of the workplace should be designed to take advantage of proper ergonomics. Each work station should be analyzed for ergonomic and motion requirements.

Ergonomics can eliminate factors in the workplace that may cause injuries and lost production. Some guidelines for providing sound ergonomic principles in the workplace include:

1. Emphasize safety at all times
2. Fit the employee to the job
3. Change the workplace to fit the employee (not vice versa)
4. Design the workplace so that neutral body positions are maintained
5. Redesign tool handles to reduce stress and injury
6. Vary the tasks through job rotation (every 2 to 4 hours)
7. Make the machine serve the human

(Sharma, 2001)[23]

Processing

Processing muda consists of additional steps or activities in the manufacturing process. Examples include the following:

- Having to remove burrs from a manufacturing process
- Reshaping a piece due to poor dies
- Adding an extra handling process due to lack of space
- Performing an inspection step (all inspection is non-value added)
- Repeating product changes that are unnecessary
- Maintaining extra copies of information

Transport

All forms of transportation are muda. This describes the use of forklifts, conveyors, pallet movers, and trucks. This can be caused by poor plant layouts, poor cell designs, use of batch processing, long lead times, large storage areas, or scheduling problems. Transport should be eliminated whenever possible.

Waste Analysis (Continued)

Waiting

The muda of waiting occurs when an operator is ready for the next operation, but must remain idle. The operator is idle due to machine downtime, lack of parts, unwarranted monitoring activities, or line stoppages. A maintenance operator waiting at a tool bin for a part is muda. The muda of waiting can be characterized by:

- Idle operators
- Breakdowns in machinery
- Long changeover times
- Uneven scheduling of work
- Batch material flows
- Long and unnecessary meetings

(Imai, 1997)[10]

Additional Classes of Wastes

In addition to the seven waste categories previously listed, some other sources of waste include:

- Misused resources
- Underutilized resources
- Counting activities
- Looking for tools or parts
- Multiple systems
- Multiple hand-offs
- Unnecessary approvals
- Machine breakdowns
- Sending bad product to customers
- Providing bad service to customers

(Metcalf, 1997)[16]

The authors have noted up to ten categories of waste from a variety of reference sources. The classes of wastes can also be differentiated between production, administrative, and transactional activities.

References

1. Anderson, P. F., & Wortman, B. L. (2012). *CQA Primer.* Terre Haute, IN: Quality Council of Indiana.

2. Bechtell, M. L. (1995). *The Management Compass: Steering the Corporation Using Hoshin Planning.* New York: AMA Membership Division.

3. Besterfield, D. H., et. al. (1999). *Total Quality Management*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

4. *Blueprints for Product Reliability: RBPR-2/RBPR-3/RBPR-4/RBPR-5/RBPR-6.* (1996). Rome, NY: Reliability Analysis Center.

5. Bowles, J. B., & Bonnell, R. D. (1994). "Failure Mode, Effects, and Criticality Analysis." 1994 Annual Reliability and Maintainability Symposium.

6. Glantz, S. A., & Slinker, B. K. (1990). *Primer of Applied Regression and Analysis of Variance.* New York: McGraw-Hill.

7. Green, S., Salkind, N., & Akey, T. (1997). *Using SPSS for Windows: Analyzing and Understanding Data.* Upper Saddle River, NJ: Prentice Hall.

8. Hashimoto, T. "Elimination of Volume Rotation Defects Through QC Circle Activities." Factory Work and QC, No. 33.

9. Huck, S. & Cormier, W. (1996). *Reading Statistics and Research*, 2nd ed. New York: HarperCollins.

10. Imai, M. (1997). *Gemba Kaizen: A Commonsense, Low-Cost Approach to Management.* New York: McGraw-Hill.

11. Ishikawa, K. (1982). *Guide to Quality Control*, White Plains: Quality Resources.

12. Ireson, W. G., Coombs, C. F., & Moss, R. (1996). *Handbook of Reliability Engineering and Management.* New York: McGraw-Hill.

13. Juran, J. M. & Gryna, F. M. (1993). *Quality Planning and Analysis*, 3rd ed. New York: McGraw-Hill.

14. Juran, J. M. (1999). *Juran's Quality Handbook*, 5th ed. New York: McGraw-Hill.

15. Kanji, G. K. (1999). *100 Statistical Tests.* London: Sage Publications Limited.

16. Metcalf, M. D. (1997). *Waste Elimination.* Sullivan, IN: Metcalf Publishing.

References (Continued)

17. MIL-STD-1629A (1980). *Procedures for Performing a Failure Mode, Effects and Criticality Analysis.*

18. Minitab. (2000). *Data Analysis and Quality Tools. Minitab User's Guide* 2. State College, PA: Minitab, Inc.

19. Porter, M. (1979, March-April). "How Competitive Forces Shape Strategy." Harvard Business Review.

20. Porter, M. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* New York: Free Press.

21. *Reliability Toolkit: Commercial Practices Edition.* (1993). Rome, New York: Reliability Analysis Center.

22. Siegel, S. (1955). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

23. Sharma, A., & Moody, P. (2001). *The Perfect Engine: How to Win in the New Demand Economy by Building to Order with Fewer Resources.* New York: The Free Press.

24. Triola, M. F., & Franklin, L. A. (1994). *Business Statistics,* Reading, MA: Addison - Wesley.

25. VBM. (2007). Information downloaded June 8, 2007 from http://www.valuebasedmanagement.net/methods_scenario_planning.html

