Statistical Inference

Making a data-based decisions is not always obvious. Consider a manufacturing process with a target of 10. The last 100 samples averaged 9.98. Should the process be adjusted to meet the target of 10? The process should be adjusted if 9.98 is statistically different than 10. If two samples of size n are taken from a population, the mean of these two samples will probably be different, and neither will be equal to the population mean. For example, a $100(1-\alpha)\%$ confidence interval contains the true population mean $100(1-\alpha)\%$ of the time. Confidence intervals and hypothesis tests are used to determine if a difference is real or if it is sampling error.

- **Confidence Intervals for the Mean**
- **Confidence Intervals for the Variance**
- **Confidence Intervals for Proportions**
- **Confidence Intervals for Rates**
- **Hypothesis Test for the Mean Against a Constant**
- **Hypothesis Test for two Means**
- **Paired Hypothesis Test for two Means**
- **Hypothesis Test for the Variance Against a Constant**
- **Hypothesis Test for two Variances**

Confidence Intervals

It is common in statistics to estimate a parameter from a sample of data. The value of the parameter using all of the possible data, not just the sample data, is called the population parameter or true value of the parameter. An estimate of the true parameter value is made using the sample data. This is called a point estimate or a sample estimate.

For example, the most commonly used measure of location is the mean. The population, or true, mean is the sum of all the members of the given population divided by the number of members in the population. As it is typically impractical to measure every member of the population, a random sample is drawn from the population. The sample mean is calculated by summing the values in the sample and dividing by the number of values in the sample. This sample mean is then used as the point estimate of the population mean.

Interval estimates expand on point estimates by incorporating the uncertainty of the point estimate. In the example for the mean above, different samples from the same population will generate different values for the sample mean. An interval estimate quantifies this uncertainty in the sample estimate by computing lower and upper values of an interval which will, with a given level of confidence contain the population parameter.

- **Confidence Intervals for the Mean**
- **Confidence Intervals for the Variance**
- **Confidence Intervals for Proportions**
- **Confidence Intervals for Rates**

Confidence Intervals for the Mean

For data that are normally distributed with unknown mean and variance, a 100(1-α)% confidence interval for the mean is

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

where: α□is the sample mean, *s* is the sample standard deviation, and $t_{\alpha/2}$ is the critical value of the *t*-distribution with ν = *n*-1 degrees of freedom.

The *t*-distribution is based on sampling from a normal distribution, but provides good results as long as the distribution is bell-shaped. Critical values of the *t*-distribution can be found using the Microsoft Excel function:

=TINV(Probability,Degrees-of-freedom,)

The normality assumption is only critical for small sample sizes. Once the sample size reaches 30 the normality assumption is not required. For data that are only slightly skewed a sample size of 3 to 4 is sufficient to ignore the normality assumption.

Example

Given the data below, construct a 90% confidence interval for the mean.

12	21	24
15	22	32

Solution

The sample mean is

$$\bar{x} = \frac{12+15+21+22+24+32}{6} = 21.0$$

The sample standard deviation is

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}} = \sqrt{\frac{6(2894) - (126)^2}{6(6-1)}} = 7.04$$

For a 90% confidence interval, α□= 0.1. From statistical tables, $t_{0.05}$ with 5 degrees of freedom is 2.015. This can also be found using the Excel formula below. A probability of 10% is used instead of 5% because the Excel function uses a two-tailed distribution.

=TINV(0.1,5)

The 90% confidence interval for the mean is

$$21 - 2.015\frac{7.04}{\sqrt{6}} < \mu < 21 + 2.015\frac{7.04}{\sqrt{6}}$$
$$15.21 < \mu < 26.79$$

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Confidence Intervals for the Variance

For data that are normally distributed with sample variance s^2, a 100(1-α)% confidence interval for the variance is

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

where, $\chi^2_{\alpha/2,n-1}$ is the critical value of the chi-square distribution with n-1 degrees of freedom.

Example

Ten boxes of cereal were randomly sampled and weighed. The sample mean was 12.1 ounces, and the sample standard deviation was 0.85. Determine a 95% confidence interval for the variance of cereal box weight.

Solution

The critical values of the chi-square distribution can be found using Microsoft Excel

$$\chi^2_{0.025,9} = 19.023$$

$$\chi^2_{0.975,9} = 2.70$$

The confidence interval is:

$$\frac{9(0.85)^2}{19.023} \le \sigma^2 \le \frac{9(0.85)^2}{2.7}$$

$$0.342 \le \sigma^2 \le 2.408$$

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Confidence Intervals for Proportions

A 100(1-α)% nonparametric confidence interval for a proportion is

$$\frac{n-r}{n-r+(r+1)F_{\alpha/2,2(r+1),2(n-r)}} \le R \le \frac{(n-r+1)F_{\alpha/2,2(n-r+1),2r}}{r+(n-r+1)F_{\alpha/2,2(n-r+1),2r}}$$

where n is the sample size and r is the number of successes.

Example

Forty items were tested with 2 failures and 38 successes. Determine an 80% confidence interval for the proportion of successes.

Solution

For a confidence level of 80%, $\alpha = 0.20$. For this example, $n = 40$ and $r = 2$. The critical values of the F-distribution can be found using the Excel formula =FINV(probability,1st degrees of freedom, 2nd degrees of freedom). For this example

$F_{0.1,6,76} = 1.853$

$F_{0.9,78,4} = 3.783$

The 80% confidence interval is

$$\frac{40-2}{40-2+(2+1)1.8531} \leq R \leq \frac{(40-2+1)3.783}{2+(40-2+1)3.783}$$

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Confidence Intervals for Rates

The chi-square distribution can be used to construct confidence intervals for the number of occurrences. Given a sample of *r* occurrences, a (1-α)% interval for the average number of occurrences is

$$\frac{X^2_{1-\alpha/2,2r}}{2} < c < \frac{X^2_{\alpha/2,2r+2}}{2}$$

Example

During a move 4 items were broken. What is the 90% confidence interval for the number of items that will be broken in a move?

Solution

For the lower confidence level, the critical chi-square value with 95% significance and 8 (4*2) degrees of freedom is required. This critical value is found using the Excel function =CHIINV(0.95,8), the critical value is 2.73. For the upper confidence level, the critical chi-square value with 5% significance and 10 (4*2+2) degrees of freedom is required. This critical value is found using the Excel function =CHIINV(0.05,10), the critical value is 18.31. The confidence interval for the number of broken items is

$2.73/2 < r < 18.31/2$

$1.37 < r < 9.315$

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Hypothesis Testing

- **Confidence Intervals**
- **Hypothesis Test for the Mean Against a Constant**
- **Hypothesis Test for two Means**
- **Paired Hypothesis Test for two Means**
- **Hypothesis Test for the Variance Against a Constant**
- **Hypothesis Test for Two Variances**

It is common in statistics to estimate a parameter from a sample of data. The value of the parameter using all of the possible data, not just the sample data, is called the population parameter or true value of the parameter. An estimate of the true parameter value is made using the sample data. This is called a point estimate or a sample estimate. For example, the most commonly used measure of location is the mean. The population, or true, mean is the sum of all the members of the given population divided by the number of members in the population. As it is typically impractical to measure every member of the population, a random sample is drawn from the population. The sample mean is calculated by summing the values in the sample and dividing by the number of values in the sample. This sample mean is then used as the point estimate of the population mean.

Interval estimates expand on point estimates by incorporating the uncertainty of the point estimate. In the example for the mean above, different samples from the same population will generate different values for the sample mean. An interval estimate quantifies this uncertainty in the sample estimate by computing lower and upper values of an interval which will, with a given level of confidence (i.e., probability), contain the population parameter.

Hypothesis Tests Hypothesis tests also address the uncertainty of the sample estimate. However, instead of providing an interval, a hypothesis test attempts to refute a specific claim about a population parameter based on the sample data. For example, the hypothesis might be one of the following:

- the population mean is equal to 10
- the population standard deviation is equal to 5
- the means from two populations are equal
- the standard deviations from 5 populations are equal

To reject a hypothesis is to conclude that it is false. However, to accept a hypothesis does not mean that it is true, only that we do not have evidence to believe otherwise. Thus hypothesis tests are usually stated in terms of both a condition that is doubted (null hypothesis) and a condition that is believed (alternative hypothesis). A common format for a hypothesis test is:

H_o:	A statement of the null hypothesis, e.g., two population means are equal.
H_a:	A statement of the alternative hypothesis, e.g., two population means are not equal.
Test Statistic:	The test statistic is based on the specific hypothesis test.

Significance Level:	The significance level, α, defines the sensitivity of the test. A value of α= 0.05 means that we inadvertently reject the null hypothesis 5% of the time when it is in fact true. This is also called the type I error. The choice of is somewhat arbitrary, although in practice values of 0.1, 0.05, and 0.01 are commonly used. The probability of rejecting the null hypothesis when it is in fact false is called the power of the test and is denoted by 1 - β. Its complement, the probability of accepting the null hypothesis when the alternative hypothesis is, in fact, true (type II error), is called and can only be computed for a specific alternative hypothesis.
Critical Region:	The critical region encompasses those values of the test statistic that lead to a rejection of the null hypothesis. Based on the distribution of the test statistic and the significance level, a cut-off value for the test statistic is computed. Values either above or below or both (depending on the direction of the test) this cut-off define the critical region.
Practical Versus Statistical Significance:	It is important to distinguish between statistical significance and practical significance. Statistical significance simply means that we reject the null hypothesis. The ability of the test to detect differences that lead to rejection of the null hypothesis depends on the sample size. For example, for a particularly large sample, the test may reject the null hypothesis that two-process means are equivalent. However, in practice the difference between the two means may be relatively small to the point of having no real engineering significance. Similarly, if the sample size is small, a difference that is large in engineering terms may not lead to rejection of the null hypothesis. The analyst should not just blindly apply the tests but should combine engineering judgment with statistical analysis.

Hypothesis Test for the Mean Against a Constant

When testing to determine if a population mean is equal to a specific value, μ_0, the null and alternative hypotheses are:

H_0: $\mu = \mu_0$
H_1: $\mu < \mu_0$, $\mu > \mu_0$, μ does not equal to μ_0

For a two-tailed test, assuming a normal sampling distribution, H_0 is rejected at significance a when the computed *t*-statistic does not fall in the range from $-t_{\alpha/2,n-1}$ to $t_{\alpha/2,n-1}$. The *t*-statistic is computed from the expression:

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

where: *n* is the sample size and *s* is the sample standard deviation.

Example

With a significance of 5%, determine if the mean of the population from which the following data was randomly sampled is statistically different than 20.

23, 28, 19, 24

Solution

With 5% significance and 3 degrees of freedom the critical values of the t-distribution are 3.18 and -3.18. These values can be found with the formula =TINV(0.05,3) in Microsoft Excel.

The average of the 4 samples is 23.5 and the sample standard deviation is 3.70. The computed *t*-statistic is 1.89. Since the computed statistic is between the 2 critical values the null hypothesis is not rejected.

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Hypothesis Test for Two Means

When testing to determine if two population means are equal, the null and alternative hypotheses are

H_0: μ_a = μ_b
H_1: $\mu_a < \mu_b$, $\mu_a > \mu_b$, μ_a does not equal μ_b

Assuming the sampling distribution is normally distributed, but the variance is unknown, the statistic tested is

$$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}}$$

For a two-tailed test, assuming a normal sampling distribution, H_0 is rejected at significance a when the computed *t*-statistic does not fall in the range from $-t_{\alpha/2,n-1}$ to $t_{\alpha/2,n-1}$. The degrees of freedom for the *t*-statistic are:

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

Example

A new beam is being designed and the design engineer wants to determine if an alloying agent affects strength. Ten beams of Design A, which includes the alloying agent, and 9 beams of design B, which does not include the alloying agent were tested to failure. The results follow. At 5% significance, can the engineer state the beams have equal strength?

Design A		Design B	
89	93	88	93
91	88	87	92
90	87	94	90
92	86	91	97
88	90	94	

Solution

There is no reason to assume the variances are equal, so the tested statistic is

$$t' = \frac{89.4 - 91.78}{\sqrt{2.22^2/10 + 3.15^2/9}} = -1.88$$

The degrees of freedom for the statistic are:

$$\nu = \frac{\left(2.22^2/10 + 3.15^2/9\right)^2}{\frac{\left(2.22^2/10\right)^2}{10-1} + \frac{\left(3.15^2/9\right)^2}{9-1}} = 14.2$$

Rounding the above value to $\nu = 14$, at a significance of 2.5% (2.5% on each tail totals 5%) the critical values of the *t*-distribution are -2.145 and 2.145. These values are found with the Excel Formula =TINV(0.05,14). The means of the two designs are accepted as equal if:

$-2.145 < t' < 2.145$

Thus, the design engineer concludes the strengths are equal (the design engineer fails to reject the null hypothesis).

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

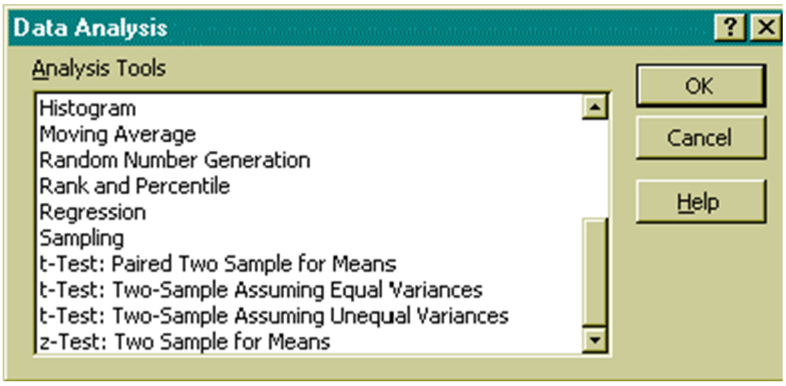
Microsoft Excel contains an *Analysis Tool Pack* that can be used to perform this test. To access this package, click the *Tools* from the menu then click *Data Analysis*. If *Data Analysis* is not on the menu the *Analysis Tool Pack* has not been activated. To activate the *Analysis Tool Pack* click the *Tools* menu and select *Add-Ins*, then check the *Analysis Tool Pack* box and click *OK*.

Solution Using Excel's Analysis Tool Pack

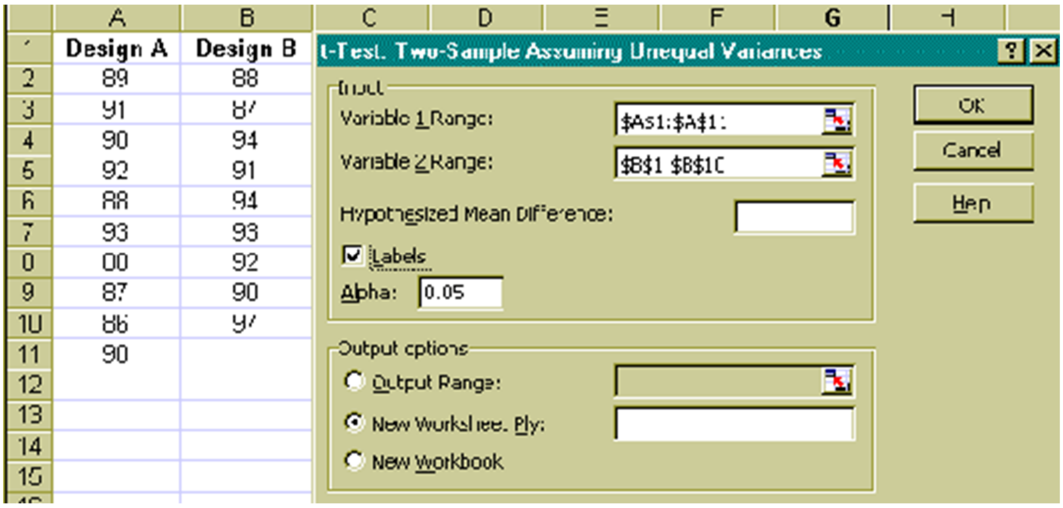
This problem is solved much easier with the assistance of Microsoft Excel. Enter the data into a spreadsheet as shown below.

	A	B
1	Design A	Design B
2	89	88
3	91	87
4	90	94
5	92	91
6	88	94
7	93	93
8	88	92
9	87	90
10	86	97
11	90	
12		

Click the "Tools" menu, select "Data Analysis" and the screen shown below will appear.



Scroll down, select "t-test: Two-Sample Assuming Unequal Variances", then click "OK". The t-test screen configuration screen will appear. Put the data for "Design A" in the "Variable 1 Range" box and the "Design B" data in the "Variable 1 Range" box. This is shown in the figure below.



Note that the data ranges include the descriptions in row1. If the descriptions are included in the data ranges, and the "Labels" box is checked, the descriptions will be shown in the output. If this is not done, the output descriptions will be "Column 1" and "Column 2". Click the "OK" button to complete the analysis. The results are shown below.

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Design A*	*Design B*
4	Mean	89.4	91.77777778
5	Variance	4.933333333	9.944444444
6	Observations	10	9
7	Hypothesized Mean Difference	0	
8	df	14	
9	t Stat	-1.880814533	
10	P(T<=t) one-tail	0.040480749	
11	t Critical one-tail	1.76130925	
12	P(T<=t) two-tail	0.080961497	
13	t Critical two-tail	2.144788596	

The null hypothesis can be evaluated by comparing the critical value from the t-distribution to the computed value, but it is easier and more informative to make decisions based on the computed significance, "P(T<=t) two-tail". The "P(T<=t) two-tail" value is equal to the probability of rejecting the null hypothesis when it is true. If this value is less than the significance of your hypothesis test, 5% in this example, the null hypothesis is rejected. If this value is greater than the significance of your hypothesis test the null hypothesis is not rejected. Since "P(T<=t) two-tail" is 8.096%, the null hypothesis is not rejected.

Paired Hypothesis Test Means

A paired test is more powerful than a standard test when comparing two populations. By paired, we mean that there is a one-to-one correspondence between the values in the two samples. That is, if $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ... , Y_n$ are the two samples, then X_i corresponds to Y_i. For paired samples, the difference $X_i - Y_i$ is calculated and the test is the same as when testing the mean against a constant. The paired test tests the difference against 0.

Example

There is concern about the ability of a new operator to properly measure the roundness of parts, so a test was conducted with the new operator and a veteran operator. Ten parts were measured, and the results are shown below. Determine if the difference in the results from the 2 operators are statistically significant.

Part	New Operator	Veteran Operator	Difference
1	1.98	1.97	0.01
2	0.42	0.40	0.02
3	0.91	0.93	-0.02
4	1.24	1.20	0.04
5	1.54	1.57	-0.03
6	1.22	1.22	0.00
7	1.47	1.46	0.01
8	1.39	1.35	0.04
9	1.81	1.82	-0.01
10	1.77	1.77	0.00

Solution

The difference is tested to determine if it is significantly different from 0. With 5% significance and 9 degrees of freedom the critical values of the t-distribution are 2.26 and -2.26. These values can be found with the formula =TINV(0.05,9) in Microsoft Excel.

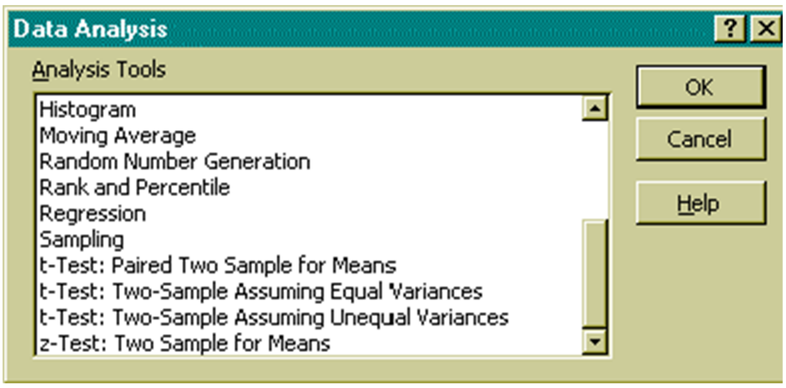
The average of the 9 differences is 0.006 and the sample standard deviation is 0.0232. The computed *t*-statistic is 0.818. Since the computed statistic is between the 2 critical values the null hypothesis is not rejected.

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Microsoft Excel contains an *Analysis Tool Pack* that can be used to perform this test. To access this package, click the *Tools* from the menu then click *Data Analysis*. If *Data Analysis* is not on the menu the *Analysis Tool Pack* has not been activated. To activate the *Analysis Tool Pack,* click File and then Options and select *Add-Ins*, then check the *Analysis Tool Pack* box and click *OK*.

Solution Using Excel's Analysis Tool Pack

This problem is solved much easier with the assistance of Microsoft Excel. Click the "Tools" menu, select "Data Analysis" and the screen shown below will appear.



Scroll down, select "t-test: Paired Two Sample for Means", then click "OK". The configuration screen will appear. Simply follow instructions.

Hypothesis Test for the Variance Against a Constant

If the population being sampled is normal, the chi-square distribution can be used to test if the variance of the distribution is equal to a specific value, σ_0. The hypotheses for this test are

H_0: $\sigma = \sigma_0$
H_1: $\sigma < \sigma_0$, $\sigma > \sigma_0$, σ does not equal to σ_0

The test statistic is

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

where:

n is the sample size,
σ is the sample standard deviation,
α is the significance, and
ν is the degrees of freedom (*n*-1).

For a two-tailed test, the null hypothesis is not rejected at a significance of α if

$$\chi^2_{1-\alpha/2,v} < \chi^2 < \chi^2_{\alpha/2,v}$$

Example

A specific type of computer monitor has a normally distributed mean life of 35,000 hours with a standard deviation of 3,000 hours. The manufacturer is testing a new design, and a random sample of 10 monitors has a standard deviation 2,800 hours. Does the new design have a lower standard deviation at 5% significance?

Solution

This is a one tailed test. The hypotheses are:

Ho: $\sigma = 3000$
H1: $\sigma < 3000$

With 9 degrees of freedom and 95% significance, the critical value of the chi-square distribution is 3.325. This is found using the Excel function =CHIINV(0.95,9).

The test statistic is

$$\chi^2 = \frac{(10-1)2800^2}{3000^2} = 7.84$$

Since the computed statistic is greater than the critical value, the null hypothesis is not rejected, and it cannot be stated that the new design has a lower variance.

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Hypothesis Test for Two Variances

When testing to determine if there is a difference in the variance of two populations, assuming the populations are normally distributed, the test statistic is

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

For a two-tailed test, the null hypothesis (the variances of the two populations are equal) is not rejected at a significance of α if

$F_{1-\alpha,v1,v2} < F < F_{\alpha,v1,v2}$

where: $\nu_1 = n_1 - 1$, and $\nu_2 = n_2 - 1$

Example

A design engineer is considering two components. Ten Type A components were tested, and the sample standard deviation of the strength was 0.092. Fifteen Type B components were tested, and the sample standard deviation of the strength was 0.097. At a significance of 0.05, is the variance of the strength of component type A different than that of component type B?

Solution

The test statistic is

$F = 0.092^2/0.097^2 = 0.8996$

The critical values of the F-distribution are:

$F_{0.925,9,14} = 0.2633$
$F_{0.025,9,14} = 3.21$

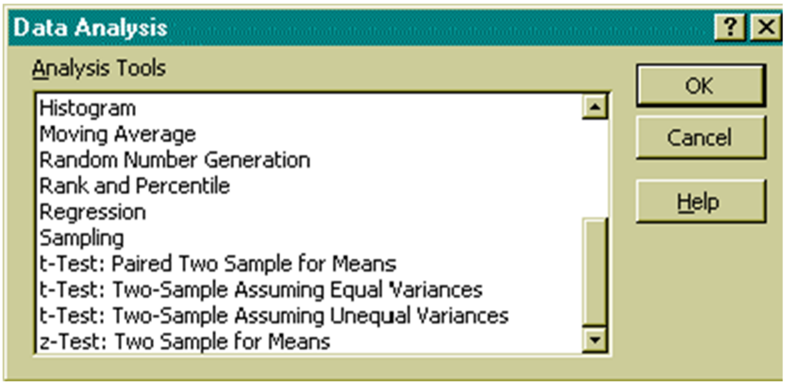
These values are found with the Excel functions =FINV(0.0275,9,14) and =FINV(0.975,9,14).

Since the calculated statistic falls between the critical values, it cannot be stated at 5% significance that the two variances are different.

Detailed calculations are available in the Microsoft Excel file "**Chapt-3 Statistical Inference.xlsx**".

Solution Using Excel's Analysis Tool Pack

This problem is solved much easier with the assistance of Microsoft Excel. Click the "Tools" menu, select "Data Analysis" and the screen shown below will appear.



Select "F-Test Two-Sample for Variance" and follow the instructions from the wizard to complete the analysis.