Equivalence and Noninferiority Tests

(Comparing Mean to Target)

Revised: 10/10/2017

Summary

This procedure tests whether the mean of a sample obtained from a single population may be considered to be equivalent to a target value. A mean is considered to be "equivalent" if it falls within a specified interval surrounding that target value. Unlike standard hypothesis tests which are designed to prove that a mean is significantly different than a specified value, equivalence tests are designed to prove that the mean is essentially equivalent to the target.

The procedure may also be used to demonstrate noninferiority. A sample is considered to be "noninferior" if the difference between the mean and the target value is no greater than (or no less than) a specified value. This situation corresponds to a one-sided test of equivalence.

Sample StatFolio: *equivalence to target.sgp*

Sample Data:

The file *process yield.sgd* contains measurements of the yield of a product produced using 3 methods (A, B, and C). 50 batches of each product were produced. A portion of the data is shown below:

C:\DocData18\process yield.sgd

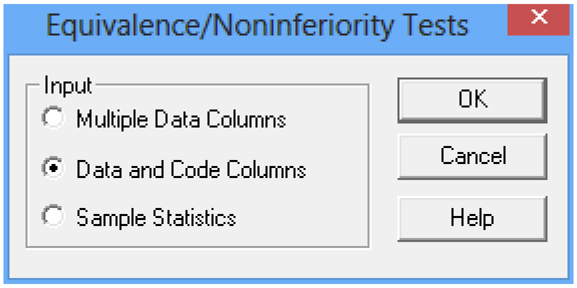
	Method	Yield	Col_
1	A	823	
2	A	739	
3	A	664	
4	A	798	
5	A	761	
6	A	807	
7	A	671	

process yield | B | C

We wish to determine which methods have means that are equivalent to a target value of 750.

Data Input

To perform the desired equivalence tests, select **Compare – Equivalence and Noninferiority Tests - Comparison to Target** from the main menu. The first dialog box displayed asks the user to specify the manner in which the data have been entered:

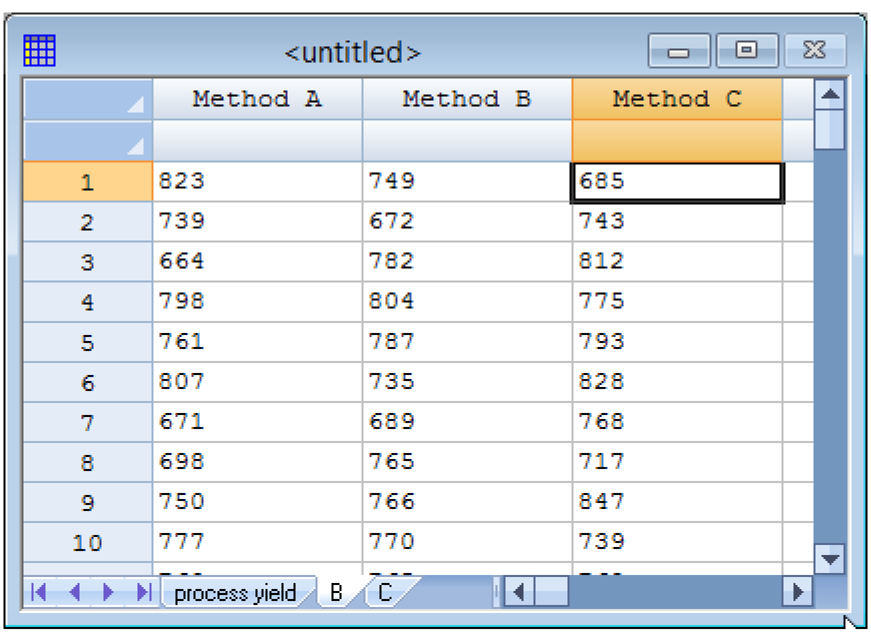


Data may be entered in any of 3 formats:

1. *Multiple Data Columns* – the data for each sample are placed in a separate column.
2. *Data and Code Columns* – all of the data are placed in a single column and a second column is created identifying the group that each observation corresponds to.
3. *Sample Statistics* – rather than entering the original data values, columns are created containing the sample sizes, means, standard deviations.

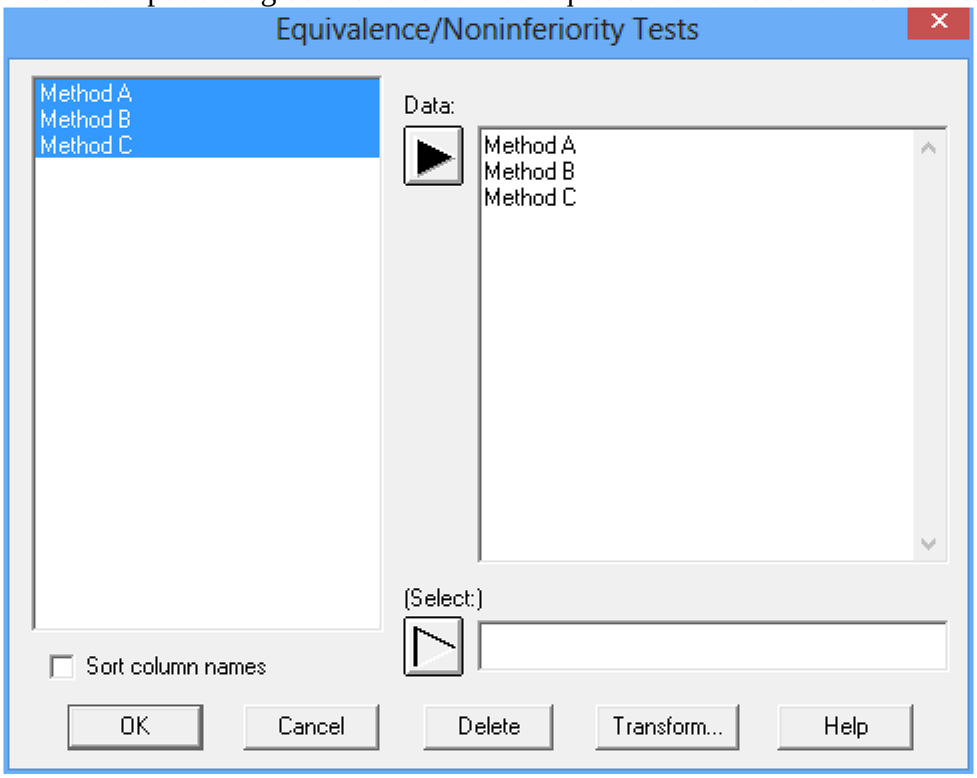
<u>Multiple Data Columns</u>

In this format, the data for each sample are placed in a separate column as in the datasheet shown below:



	Method A	Method B	Method C
1	823	749	685
2	739	672	743
3	664	782	812
4	798	804	775
5	761	787	793
6	807	735	828
7	671	689	768
8	698	765	717
9	750	766	847
10	777	770	739

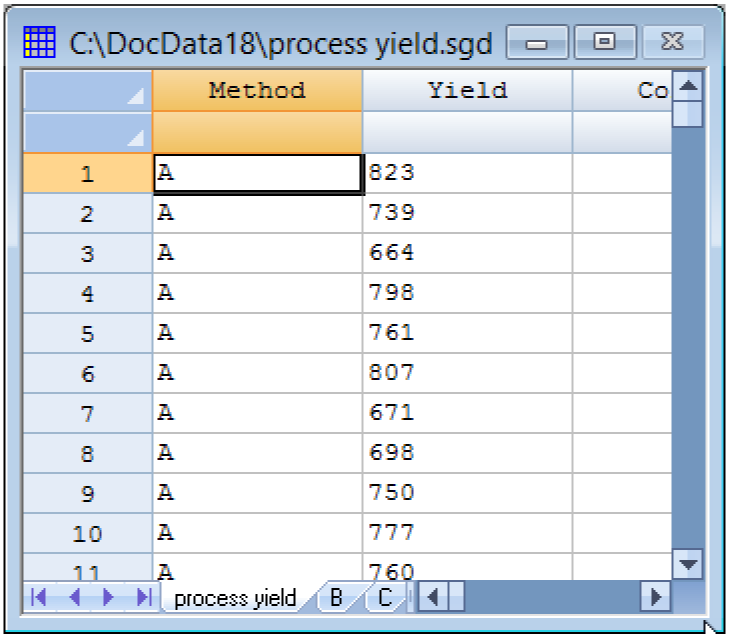
The data input dialog box for this format requests the names of the columns:



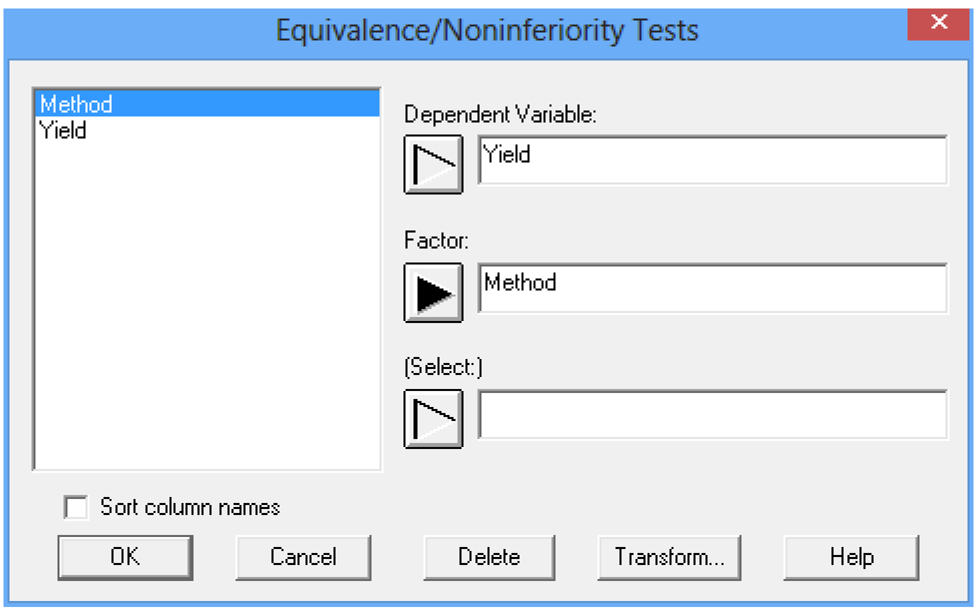
At least 2 columns containing data must be specified.

<u>Data and Code Columns</u>

In this format, all of the data is placed in a single column and a second column is created identifying which data belong to which samples:



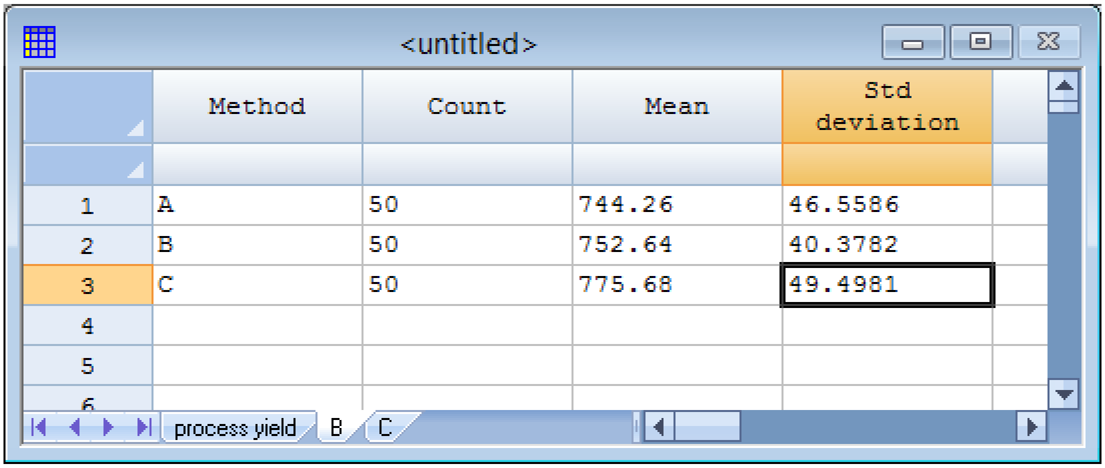
The data input dialog box for this format requests the names of the data and code columns:



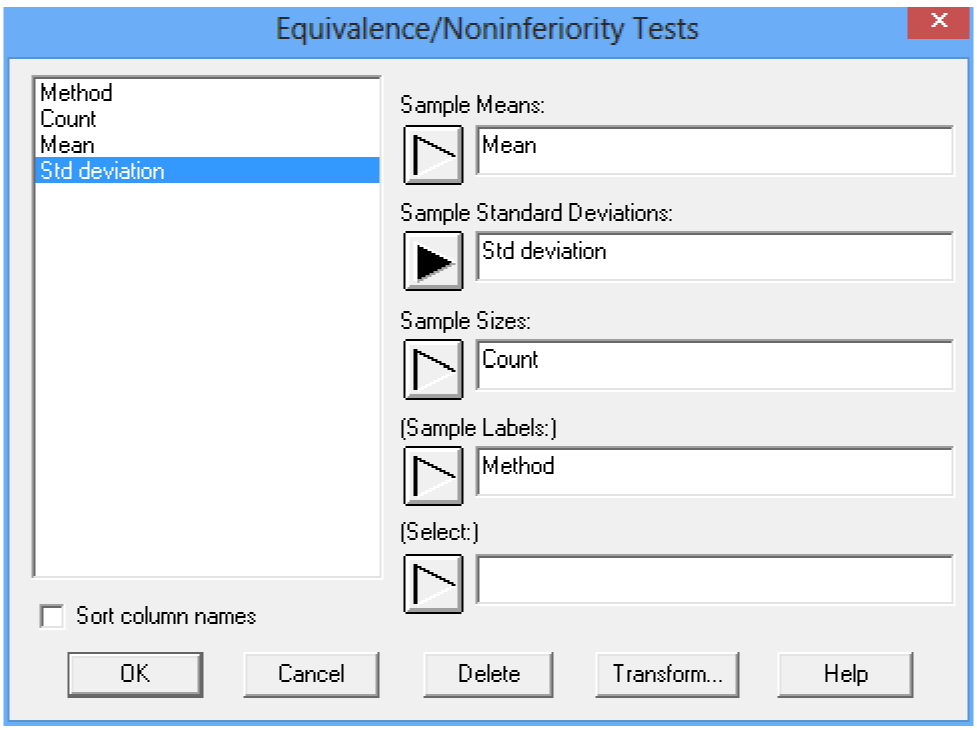
Note that the *Factor* column may be either numeric or character.

<u>Sample Statistics</u>

In this format, columns are created to contain the sample sizes, means, standard deviations and an optional label for each sample:

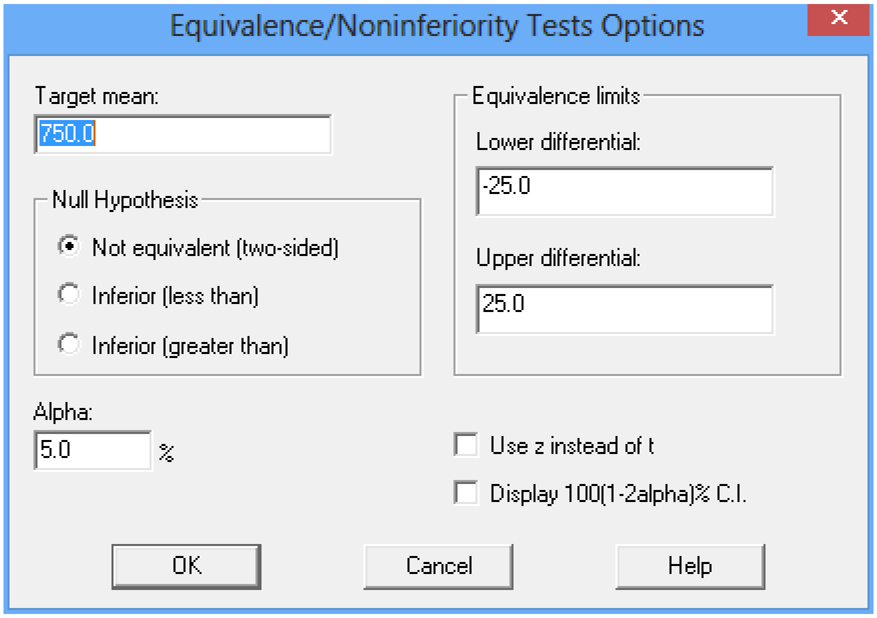


The data input dialog box for this format requests the names of the columns with the sample statistics:



Analysis Options

Once the data is specified, a third dialog box is displayed on which to specify the hypothesis to be tested.



The most common type of test is a two-sided test of equivalence. In such a test, the null hypothesis is that the mean of the population from which the sample has been obtained, μ, is not equivalent to a target value T. By not equivalent, we mean that the difference between the mean and the target μ - T is either less than some lower differential Δ_L, or greater than some upper differential Δ_U:

Null hypothesis: $\mu - T < \Delta_L$ or $\mu - T > \Delta_U$

If this hypothesis is rejected, then we will have demonstrated that the difference between the mean and the target satisfies $\Delta_L \leq \mu - T \leq \Delta_U$, which is our definition of equivalence.

To demonstrate equivalence, Statgraphics uses the TOST procedure of Schuirman (1987). This procedure consists of two one-sided tests: an upper-tailed test used to demonstrate that $\mu - T \geq \Delta_L$ and a lower-tailed test used to demonstrate that $\mu - T \leq \Delta_U$. Obtaining significant results on both tests allows an assertion of equivalence between the means.

The fields on the *Analysis Options* dialog box specify:

- **Target mean:** the target value T.

- **Null hypothesis:** whether to perform a two-tailed test of equivalence as described above or a one-tailed test of noninferiority. In the latter case, the null hypothesis is one of the following:

  "Less than" null hypothesis: $\mu - T < \Delta_L$

  "Greater than" null hypothesis: $\mu - T > \Delta_U$

- **Equivalence limits:** the value of the lower differential Δ_L and the upper differential Δ_U.

- **Alpha:** the significance level at which the tests will be performed.
- **Use z instead of t:** whether to perform a z-test rather than a t-test.
- **Display 100(1-2alpha)% C.I.:** when displaying confidence intervals, use (1-2α) instead of (1-α).

Analysis Summary

The *Analysis Summary* for the sample data using the default options is shown below:

Equivalence/Noninferiority Tests - Comparison to Target

Dependent variable: Yield
Factor: Method

Sample Statistics

Sample	*n*	*Minimum*	*Maximum*	*Mean*	*Std. deviation*
A	50	664.0	844.0	744.26	46.5586
B	50	672.0	844.0	752.64	40.3782
C	50	667.0	892.0	775.68	49.4981

Equivalence Analysis

Target mean: 750.0
Null hypothesis: Not equivalent (two-sided)
Lower equivalence differential: -25.0
Upper equivalence differential: 25.0

Comparison	*Difference*	*Stnd. error*	*Lower 90% CL*	*Upper 90% CL*
A v target	-5.74	6.58439	733.221	755.299
B v target	2.64	5.71034	743.066	762.214
C v target	25.68	7.00008	750.0	787.416

Comparison	Lower t-value	Upper t-value	Lower P-value	Upper P-value
A v target	2.9251	-4.66862	0.00260204	0.0000120654
B v target	4.84034	-3.9157	0.0000067789	0.000139494
C v target	7.23991	0.0971417	1.41071E-9	0.538495

Comparison	Maximum P-value	Conclusion (alpha=5%)
A v target	0.00260204	Equivalence has been demonstrated.
B v target	0.000139494	Equivalence has been demonstrated.
C v target	0.538495	Equivalence has not been demonstrated.

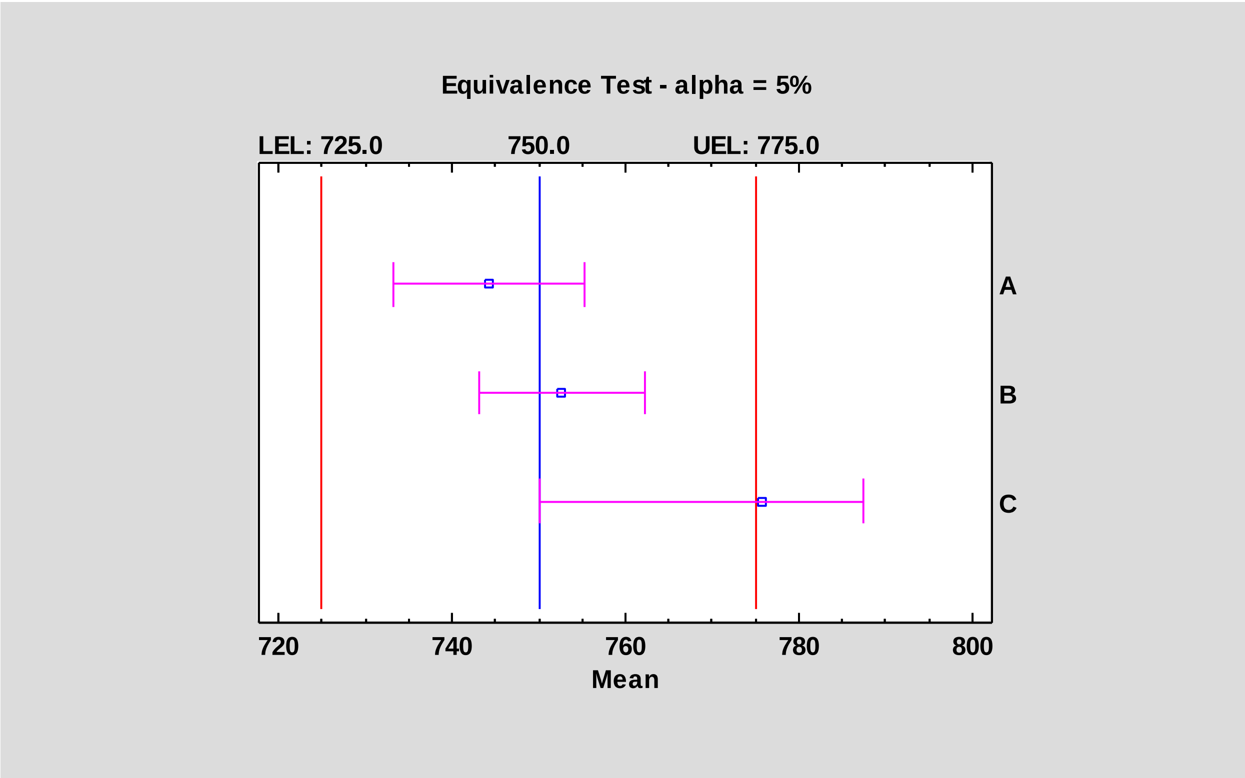
The top of the output displays summary statistics for each sample. This is followed by an *Equivalence Analysis* which compares each sample mean to the target value. In the example, the null hypothesis is that the mean is not within the equivalence range of 725 to 775 calculated from 750±25.

The output then displays the estimated difference between each mean and the target, together with a 100(1-2α)% conference interval for the mean. If the confidence interval is entirely within the equivalence range, then equivalence can be asserted. Otherwise, it cannot. In the example, only methods A and B have both confidence limits between 725 and 775.

An equivalent method for determining whether a mean is equivalent to a target value is to run two one-sided tests, one against the target value plus the lower differential and another against the target value plus the upper differential. If both P-values are less than α, then equivalence can be asserted. The summary table shows the greater of the two P-values for each mean and asserts equivalence only for methods A and B.

Equivalence Plot

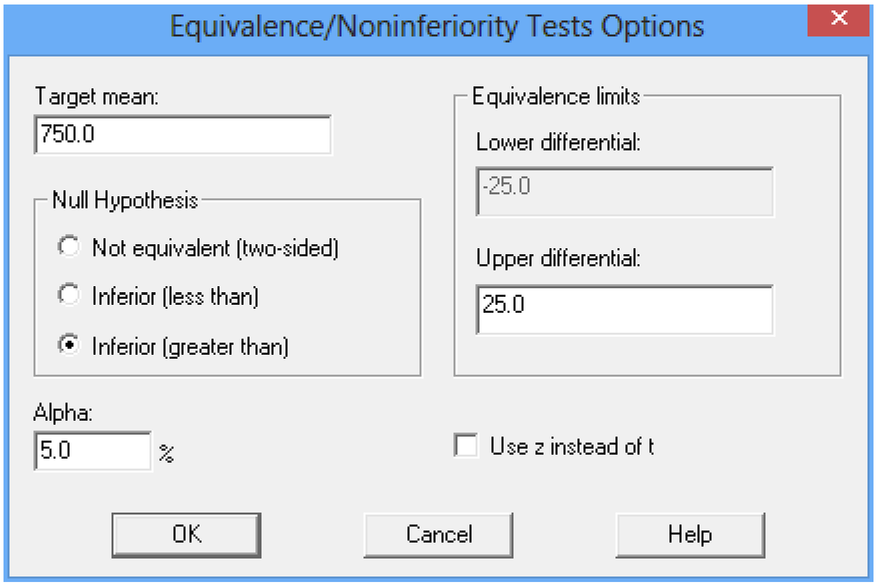
This plot shows the confidence intervals for each mean. If an interval is contained entirely in the region between the lower and upper equivalence limits, then the means may be asserted to be equivalent.



One-Sided Noninferiority Tests

In some circumstances, the desired goal is not one of showing that the difference between a mean and a target value is within some specified range. Instead, the goal is either to show that the difference is no bigger than some value Δ_U or to show that the difference is no smaller than some value Δ_L. Rejection of a null hypothesis in such a one-sided situation leads to the assertion that the mean is not inferior to the target (it might be either equivalent or superior).

For example, suppose it was desired to show that the mean was no more than 25 units more than the target. In such a case, the *Analysis Options* dialog box would be completed as shown below:



In this case, the null hypothesis is $\mu - T > 25$. If this hypothesis can be rejected, then we can claim that the mean is not inferior to the target.

For the sample data, the *Analysis Summary* is shown below:

Equivalence/Noninferiority Tests - Comparison to Target

Dependent variable: Yield
Factor: Method

Sample Statistics

Sample	*n*	*Minimum*	*Maximum*	*Mean*	*Std. deviation*
A	50	664.0	844.0	744.26	46.5586
B	50	672.0	844.0	752.64	40.3782
C	50	667.0	892.0	775.68	49.4981

Equivalence Analysis

Target mean: 750.0
Null hypothesis: Inferior (greater than)
Upper equivalence differential: 25.0

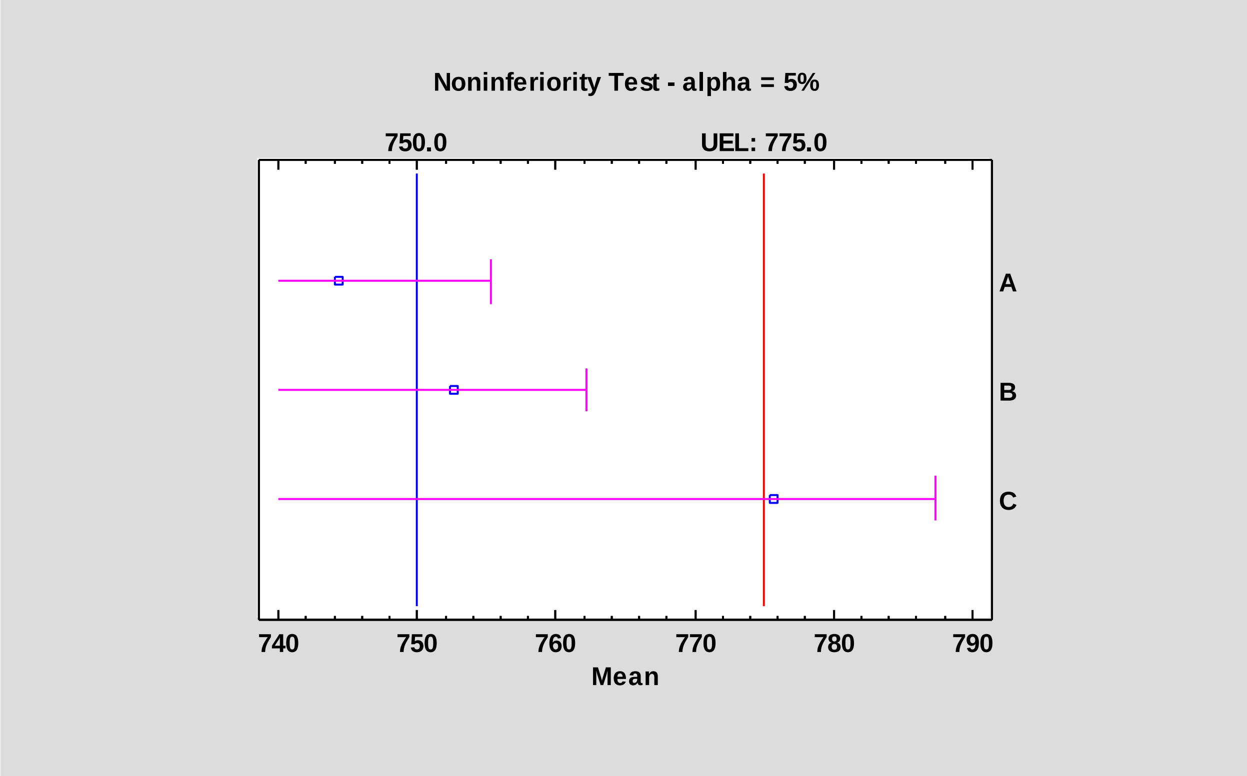
Comparison	*Difference*	*Stnd. error*	*Upper 95% CL*
A v target	-5.74	6.58439	755.299
B v target	2.64	5.71034	762.214
C v target	25.68	7.00008	787.416

Comparison	Upper t-value	Upper P-value
A v target	-4.66862	0.0000120654
B v target	-3.9157	0.000139494
C v target	0.0971417	0.538495

Comparison	Maximum P-value	Conclusion (alpha=5%)
A v target	0.0000120654	Noninferiority has been demonstrated.
B v target	0.000139494	Noninferiority has been demonstrated.
C v target	0.538495	Noninferiority has not been demonstrated.

For each sample, the output displays an upper confidence bound for the mean. If the upper confidence bound is less than the target plus the upper equivalence differential, the P-value of a lower-tailed test comparing the difference to Δ_U will be less than alpha and noninferiority may be asserted.

The *Equivalence Plot* displays the one-sided confidence bounds for each mean:



Noninferiority may be asserted for any mean in which the confidence bound does not contain the UEL.

Calculations

By default, the confidence intervals are calculated by:

$$\left[min\left(T\,, \bar{x} - t_{\alpha,\nu} s \sqrt{\frac{1}{n}} \right), max\left(T\,, \bar{x} + t_{\alpha,\nu} s \sqrt{\frac{1}{n}} \right) \right]$$

If "Display 100(1-2alpha) C.I." is selected on the Analysis Options dialog box, the confidence intervals are calculated by:

$$\bar{x} \pm t_{\alpha,\nu} s \sqrt{\frac{1}{n}}$$

References

Berger, R.L. and Hsu, J.C. (1995). "Bioequivalence trials, intersection-union tests, and equivalence confidence sets." Institute of Statistics Mimeo Series Number 2279.

Chow, S.-H. and Shao, J. (2002). Statistics in Drug Research: Methodologies and Recent Developments. New York: Marcel-Dekker.

Hsu, J.C., Hwang, J.T.G., Liu, H.-K., and Ruberg, S.J. (1994). "Confidence intervals associated with tests for bioequivalence." Biometrika 81: 103-114.

Schuirmann, D.J. (1987). "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability." J. Pharmacokinet Biopharm. 15(6): 657-680.