Equivalence and Noninferiority Tests (Comparing Paired Samples)

Revised: 10/10/2017

Summary 1
Data Input 3
Analysis Options 4
Analysis Summary 5
Equivalence Plot 6
One-Sided Noninferiority Tests 7
Calculations 9
References 9

Summary

This procedure tests whether the means of 2 samples may be considered equivalent, assuming that the data in the 2 samples consist of matched pairs. Two samples are considered to be “equivalent” if the difference between their respective means falls within some specified interval surrounding 0. Unlike standard hypothesis tests which are designed to prove superiority of one method over another, equivalence tests are designed to prove that two methods have essentially the same mean.

The procedure may also be used to demonstrate noninferiority. Two samples are considered to be “noninferior” if the difference the mean of one sample and the mean of another is no greater than (or no less than) a specified value. This situation corresponds to a one-sided test of equivalence.

Sample StatFolio: *equivalence for paired data.sgp*

Sample Data:

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

C:\DocData18\process yield in columns.sgd

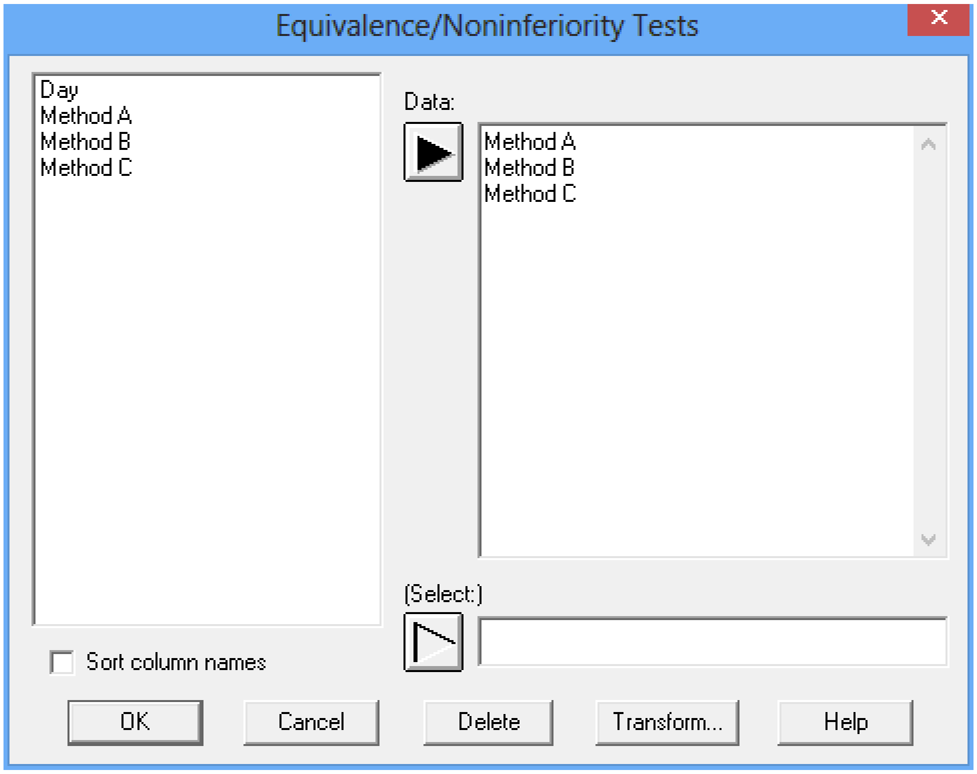
	Day	Method A	Method B	Method C
1	1	823	749	685
2	2	739	672	743
3	3	664	782	812
4	4	798	804	775
5	5	761	787	793
6	6	807	735	828
7	7	671	689	768
8	8	698	765	717

process yield in columns | B | C

We wish to demonstrate that the 3 methods produce equivalent yields, where any 2 methods are considered to be equivalent if their mean yields differ by no more than 25.

Data Input

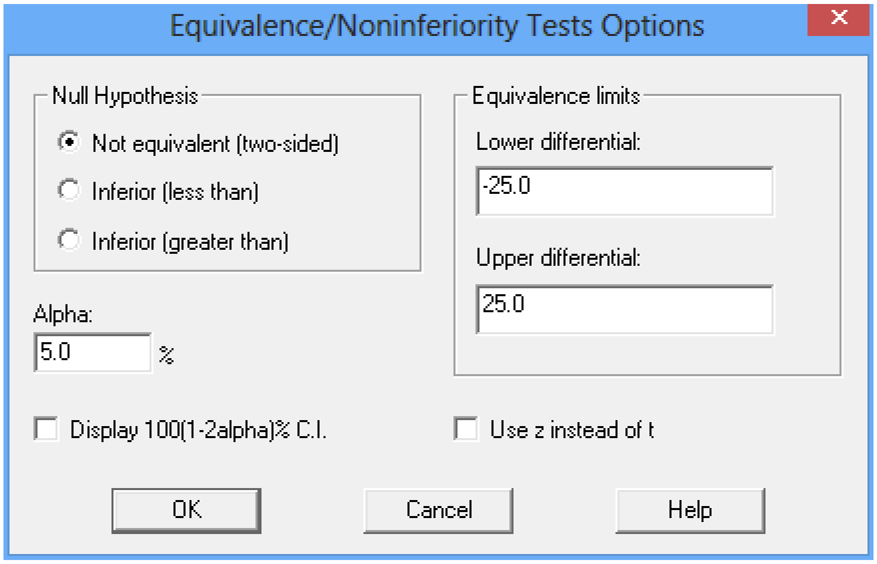
Since each row consists of measurements made on a particular day, the data in any 2 columns form matched pairs. To perform the desired equivalence tests, select **Compare - Equivalence and Noninferiority Tests - Comparison of Paired Samples** from the main menu. The data input dialog box requests the names of the columns containing the samples to be compared:



At least 2 columns containing data must be specified. If more than 2 columns are specified, a comparison will be made for each possible pair.

Analysis Options

Once the data is specified, a second 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 means of the two samples being compared, μ_1 and μ_2, are not equivalent. By not equivalent, we mean that the difference between the means μ_1-μ_2 is either less than some lower differential Δ_L, or greater than some upper differential Δ_U:

Null hypothesis: $\mu_1 - \mu_2 < \Delta_L$ or $\mu_1 - \mu_2 > \Delta_U$

If this hypothesis is rejected, then we will have demonstrated that the difference between the means satisfies $\Delta_L \leq \mu_1 - \mu_2 \leq \Delta_U$, 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 $\mu_1 - \mu_2 \geq \Delta_L$ and a lower-tailed test used to demonstrate that $\mu_1 - \mu_2 \leq \Delta_U$. Obtaining significant results on both tests allows an assertion of equivalence between the means.

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

- **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: $\mu_1 - \mu_2 < \Delta_L$

  "Greater than" null hypothesis: $\mu_1 - \mu_2 > \Delta_U$

- **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:** requests that a paired z-test be performed rather than a paired 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 of Paired Samples
Sample 1: Method A
Sample 2: Method B
Sample 3: Method C

Sample Statistics

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

Equivalence Analysis
Null hypothesis: Not equivalent (two-sided)
Lower equivalence differential: -25.0
Upper equivalence differential: 25.0

Comparison	*n*	*Difference*	*Stnd. error*	*Lower 90% CL*	*Upper 90% CL*
Method A v Method B	50	-8.38	8.78444	-23.1076	6.3476
Method A v Method C	50	-31.42	9.15245	-46.7646	0.0
Method B v Method C	50	-23.04	9.38588	-38.7759	0.0

Comparison	Lower t-value	Upper t-value	Lower P-value	Upper P-value
Method A v Method B	1.89198	-3.7999	0.0322	0.0002
Method A v Method C	-0.701451	-6.16447	0.7568	0.0000
Method B v Method C	0.208824	-5.11833	0.4177	0.0000

Comparison	Maximum P-value	Conclusion (alpha=5%)
Method A v Method B	0.0322	Equivalence has been demonstrated.
Method A v Method C	0.7568	Equivalence has not been demonstrated.
Method B v Method C	0.4177	Equivalence has not been demonstrated.

The top of the output displays summary statistics for each sample. These statistics are calculated using all available data in each column. This is followed by an *Equivalence Analysis* which compares each pair of sample means. In the example, the null hypothesis is that the difference between the means is not within the equivalence range of -25 to 25.

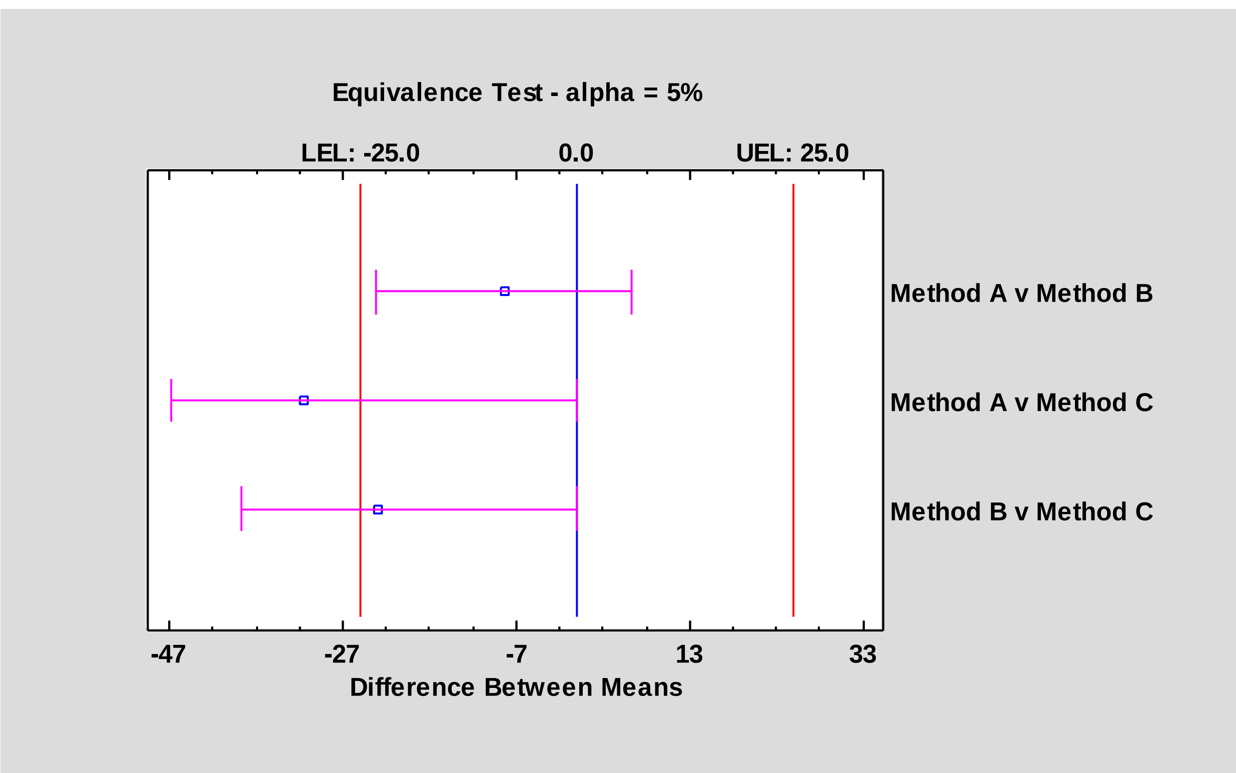
The output then displays the estimated difference between each pair of means, together with a 100(1-2α)% conference interval for the difference. If the confidence interval is entirely within the equivalence range, then equivalence can be asserted. Otherwise, it cannot. In the example, only methods A and B have both confidence limits between -25 and 25.

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

Note: When comparing two samples, the difference between the means is calculated using all rows in the datasheet that have non-missing values for those two samples. If missing data is present, the calculated difference may not equal the difference between the means displayed in the summary statistics table.

Equivalence Plot

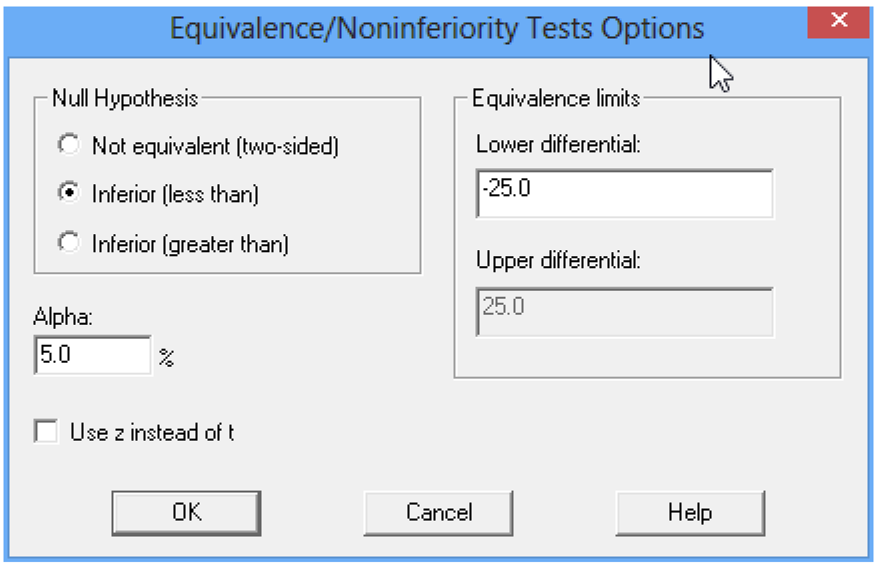
This plot shows the confidence intervals for each pair of means. If an interval is contained entirely in the region between the lower and upper equivalence limits, then the means may be asserted to be equivalent.



One-Sided Noninferiority Tests

In some circumstances, the desired goal is not one of showing that the difference between 2 means 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 one mean is not inferior to another mean (it might be either equivalent or superior).

For example, suppose it was desired to show that the mean of method 1 was no more than 25 units less than the mean of method 2. In such a case, the *Analysis Options* dialog box would be completed as shown below:



In this case, the null hypothesis is $\mu_1 - \mu_2 < -25$. If this hypothesis can be rejected, then we can claim that method 1 is not inferior to method 2.

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

Equivalence/Noninferiority Tests - Comparison of Paired Samples

Sample 1: Method A
Sample 2: Method B
Sample 3: Method C

Sample Statistics

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

Equivalence Analysis
Null hypothesis: Inferior (less than)
Lower equivalence differential: -25.0

Comparison	*n*	*Difference*	*Stnd. error*	*Lower 95% CL*
Method A v Method B	50	-8.38	8.78444	-23.1076
Method A v Method C	50	-31.42	9.15245	-46.7646
Method B v Method C	50	-23.04	9.38588	-38.7759

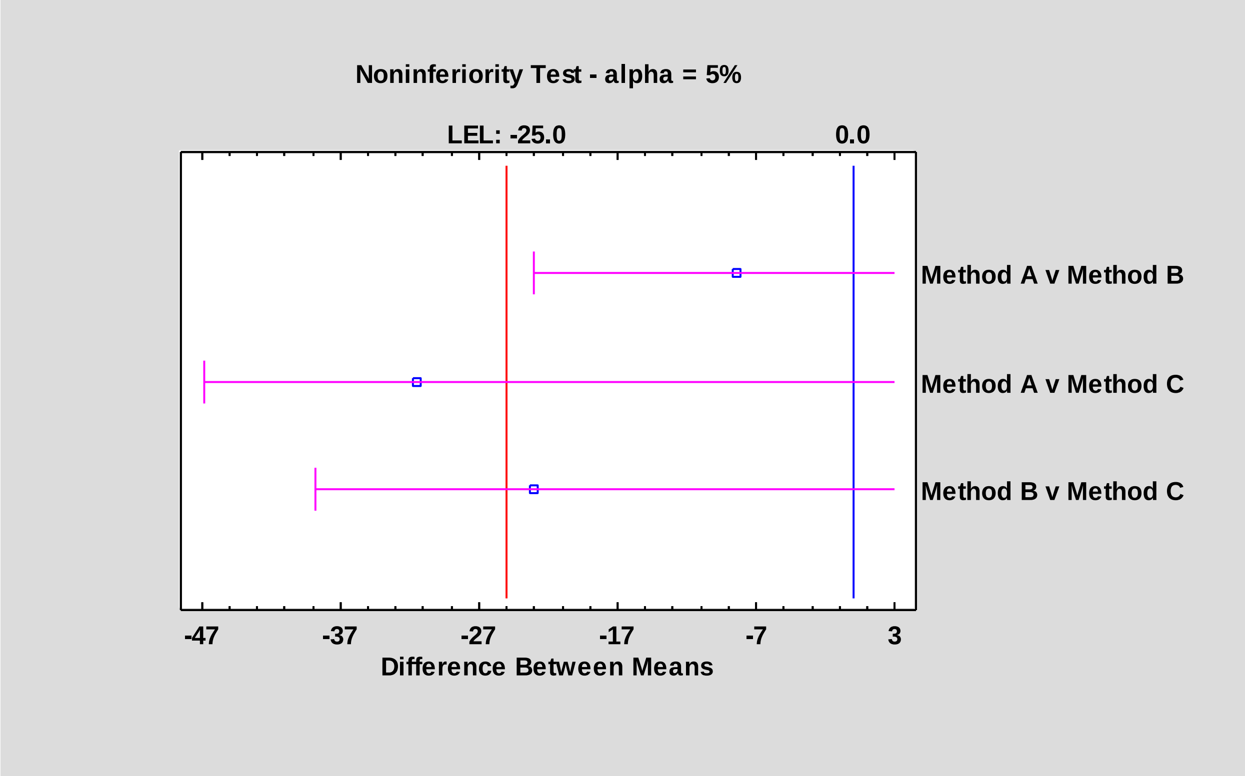
Comparison	Lower t-value	Lower P-value
Method A v Method B	1.89198	0.0322
Method A v Method C	-0.701451	0.7568
Method B v Method C	0.208824	0.4177

Comparison	Maximum P-value	Conclusion (alpha=5%)
Method A v Method B	0.0322	Noninferiority has been demonstrated.
Method A v Method C	0.7568	Noninferiority has not been demonstrated.
Method B v Method C	0.4177	Noninferiority has not been demonstrated.

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

NOTE: The order in which the samples are entered is important in this case, since the null hypothesis is that the first sample in each comparison is inferior to the second. Be sure to enter your samples in whatever order gives you the test you desire.

The *Equivalence Plot* displays the one-sided confidence bounds for the difference between each pair of means:



Noninferiority may be asserted for any differences in which the confidence bounds do not contain the LEL.

Calculations

By default, the confidence intervals are calculated by:

$$\left[min\left(0\,,\bar{x}_1-\bar{x}_2-t_{\alpha,\nu}s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right),max\left(0\,,\bar{x}_1-\bar{x}_2+t_{\alpha,\nu}s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)\right]$$

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

$$\bar{x}_1-\bar{x}_2\pm t_{\alpha,\nu}s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

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.