Equivalence and Noninferiority Tests (Comparing Two Variances)

statgraphics®
centurion

Revised: 8/11/2019

Summary

This procedure tests whether the variances or standard deviations of 2 samples may be considered equivalent. Two samples are considered to be "equivalent" if the ratio of their respective standard deviations falls within some specified interval surrounding 1. 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 variance.

The procedure may also be used to demonstrate noninferiority. A sample is considered to be "noninferior" compared to another sample if the ratio of their respective standard deviations is no greater than (or no less than) a specified value. This situation corresponds to a one-sided test of equivalence.

Sample StatFolio: *equivsigma.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:
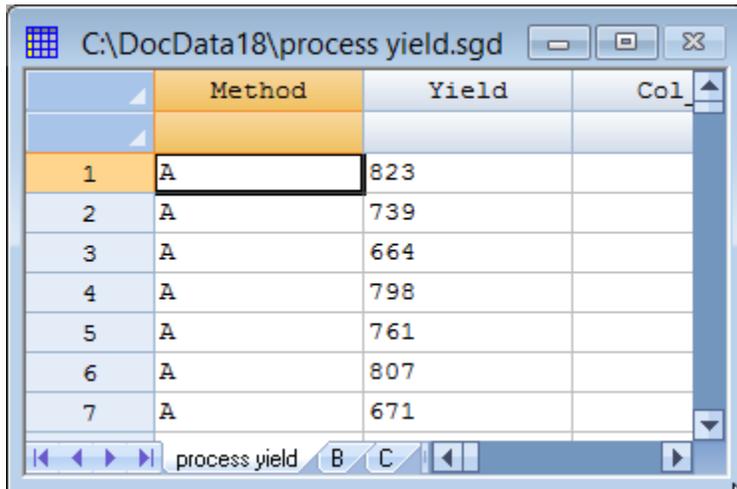

We wish to demonstrate that the 3 methods produce equivalent yields, where any 2 methods are considered to be equivalent if their standard deviations differ by no more than 20%.

Data Input

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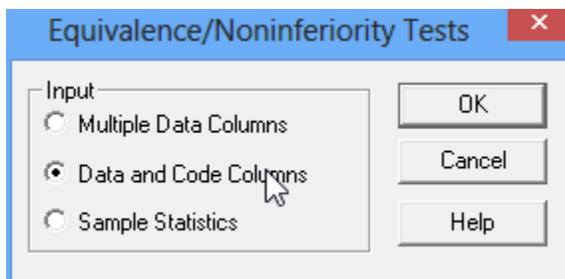

Data may be entered in any of 3 formats:

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.

Multiple Data Columns

In this format, the data for each sample are placed in a separate column as in the datasheet shown below:
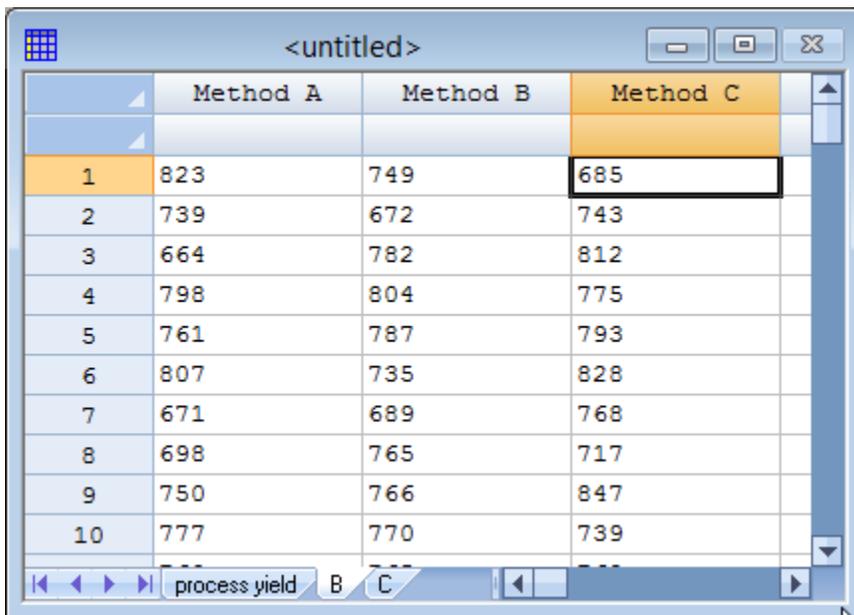
	Method A	Method B	Method C
1	823	749	685
2	739	672	743
3	664	782	812
4	798	804	775
5	761	787	793
6	807	735	828
7	671	689	768
8	698	765	717
9	750	766	847
10	777	770	739

The data input dialog box for this format requests the names of the columns:
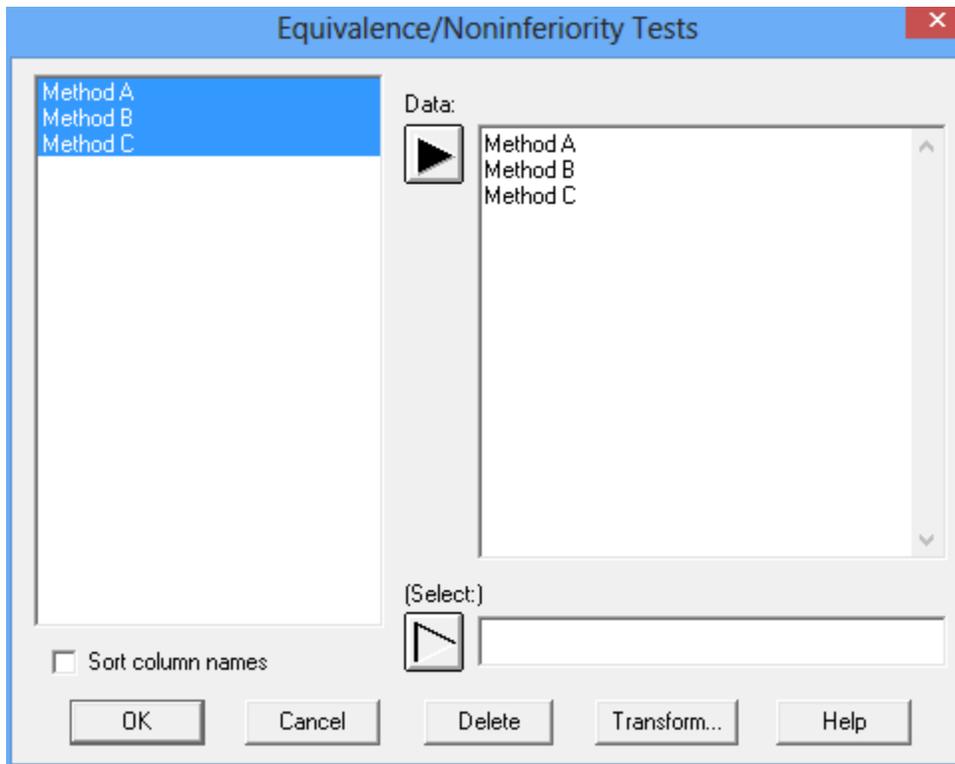
At least 2 columns containing data must be specified.

Data and Code Columns

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.

Sample Statistics

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

```
Equivalence/Noninferiority Tests Options                    ×

┌─ Null Hypothesis ──────────────┐   ┌─ Equivalence limits ──────────┐
│  ⦿ Not equivalent (two-sided)  │   │  Lower ratio:                 │
│                                │   │  ┌──────────────────────────┐ │
│  ○ Inferior (less than)        │   │  │ 0.8                      │ │
│                                │   │  └──────────────────────────┘ │
│  ○ Inferior (greater than)     │   │                               │
│                                │   │  Upper ratio:                 │
└────────────────────────────────┘   │  ┌──────────────────────────┐ │
                                      │  │ 1.2                      │ │
Alpha:                                │  └──────────────────────────┘ │
┌──────────────┐                      │                               │
│ 5.0          │ %                    └───────────────────────────────┘
└──────────────┘

☐ Display 100(1-2alpha)% C.I.


    ┌─────────┐      ┌─────────┐      ┌─────────┐
    │   OK    │      │ Cancel  │      │  Help   │
    └─────────┘      └─────────┘      └─────────┘
```
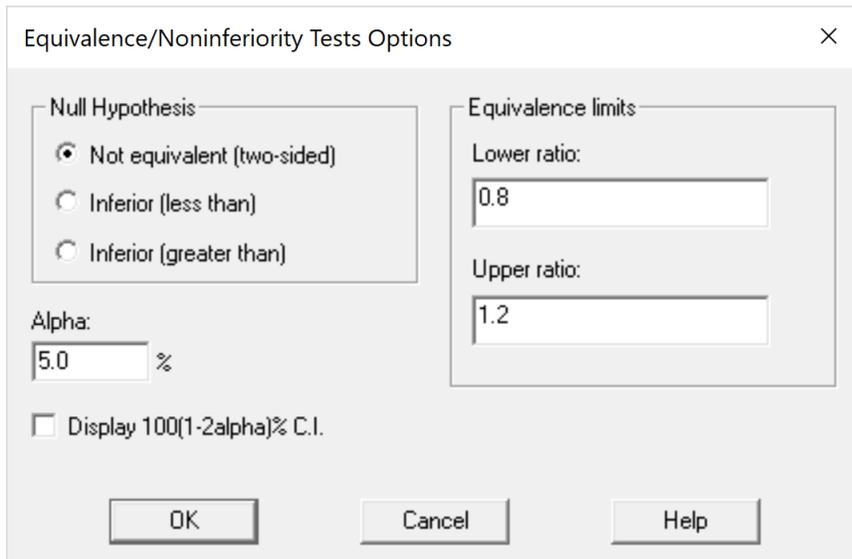
The most common type of test is a two-sided test of equivalence. In such a test, the null hypothesis is that the standard deviations of the two samples being compared, σ_1 and σ_2, are not equivalent. By not equivalent, we mean that the ratio of the standard deviations σ_1/σ_2 is either less than some lower differential Δ_L, or greater than some upper differential Δ_U:

$$\text{Null hypothesis: } \sigma_1 / \sigma_2 < \Delta_L \text{ or } \sigma_1 / \sigma_2 > \Delta_U$$

If this hypothesis is rejected, then we will have demonstrated that the ratio of the standard deviations satisfies $\Delta_L \leq \sigma_1 / \sigma_2 \leq \Delta_U$, 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 $\sigma_1 / \sigma_2 \geq \Delta_L$ and a lower-tailed test used to demonstrate that $\sigma_1 / \sigma_2 \leq \Delta_U$. Obtaining significant results on both tests allows an assertion of equivalence between the standard deviations.

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:

    $$\text{"Less than" null hypothesis: } \sigma_1 / \sigma_2 < \Delta_L$$

    $$\text{"Greater than" null hypothesis: } \sigma_1 / \sigma_2 > \Delta_U$$

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

- **Alpha:** the significance level at which the tests will be performed.

- **Display 100(1-2alpha)% C.I.:** when displaying confidence intervals, use $(1-2\alpha)$ instead of $(1-\alpha)$.

Analysis Summary

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

Equivalence & Noninferiority Tests - Comparison of Variances for Two Independent Samples

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
Null hypothesis: Not equivalent (two-sided)
Lower equivalence ratio: 0.8
Upper equivalence ratio: 1.2

Comparison	Ratio	95% CL	95% CL
A v B	1.15306	0.909508	1.46184
A v C	0.940615	0.741933	1.1925
B v C	0.815753	0.643445	1.0342

Comparison	Lower F-value	Upper F-value	Lower P-value	Upper P-value
A v B	2.07743	0.923303	0.0059	0.3906
A v C	1.38243	0.614415	0.1302	0.0457
B v C	1.03977	0.46212	0.4460	0.0040

| Comparison	Maximum P-value	Conclusion (alpha=5%)
A v B	0.3906	Equivalence has not been demonstrated.
A v C	0.1302	Equivalence has not been demonstrated.
B v C	0.4460	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 pair of sample standard deviations. In the example, the null hypothesis is that the ratio of the standard deviations is not within the equivalence range of 0.8 to 1.2.

The output then displays the estimated ratio of each pair of standard deviations, together with a $100(1-2\alpha)\%$ conference interval for the ratio. If the confidence interval is entirely within the equivalence range, then equivalence can be asserted. Otherwise, it cannot. In the example, no pair of methods have both confidence limits between 0.8 and 1.2.

An equivalent method for determining whether two standard deviations 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 than α, then equivalence can be asserted. The summary table shows the greater of the two P-values for each pair of standard deviations and cannot assert equivalence for any two methods.
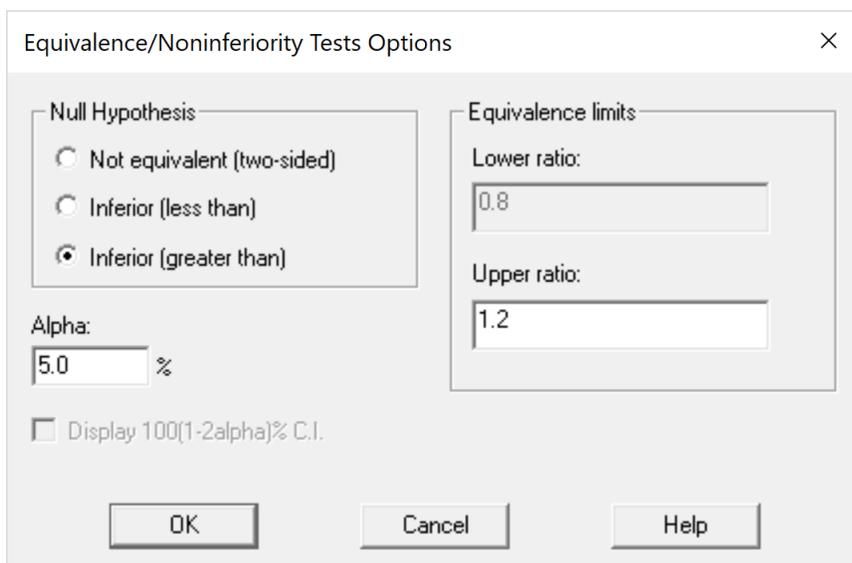
Equivalence Plot

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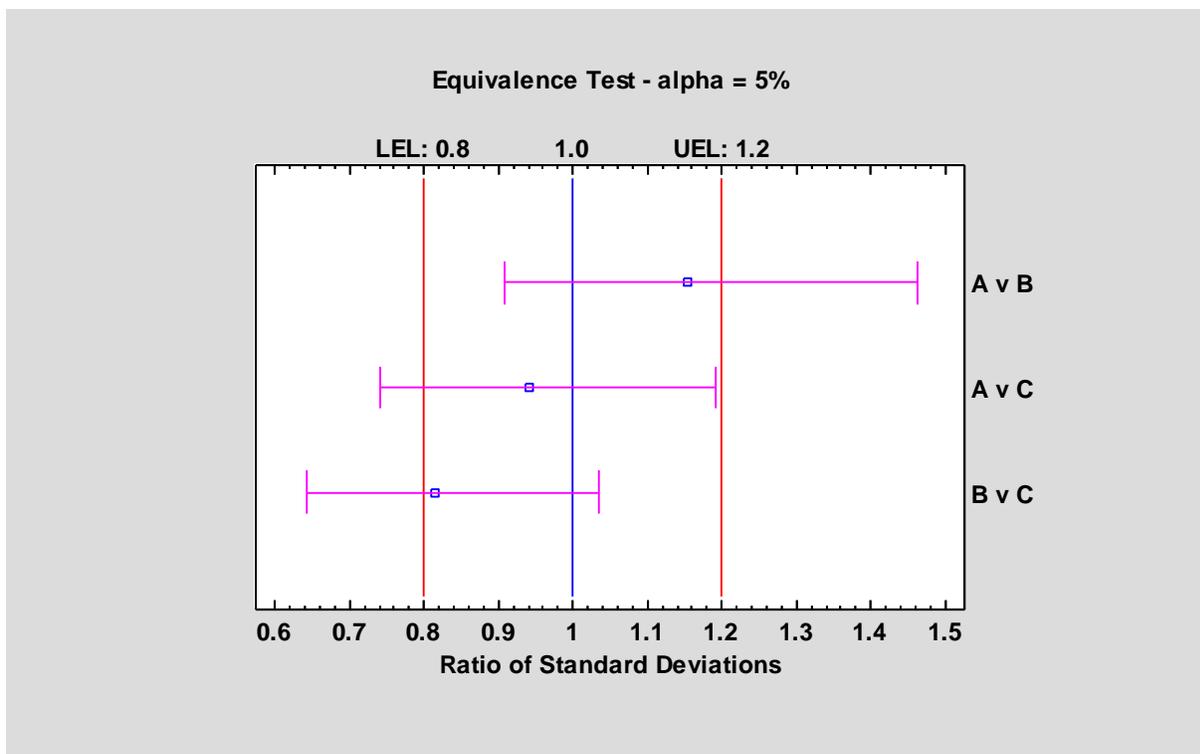
One-Sided Noninferiority Tests

In some circumstances, the desired goal is not one of showing that the ratio of 2 standard deviations is within some specified range. Instead, the goal is either to show that the ratio is no bigger than some value Δ_U or to show that the ratio 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 standard deviation is not inferior to another standard deviation (it might be either equivalent or superior).

For example, suppose it was desired to show that the standard deviation of method 1 was no more than 1.2 times the standard deviation 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 $\sigma_1 / \sigma_2 > 1.2$. 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 Variances for Two Independent Samples

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

Null hypothesis: Inferior (greater than)

Upper equivalence ratio: 1.2

Comparison	Ratio	95% CL
A v B	1.15306	1.46184
A v C	0.940615	1.1925
B v C	0.815753	1.0342

Comparison	Upper F-value	Upper P-value
A v B	0.923303	0.3906
A v C	0.614415	0.0457
B v C	0.46212	0.0040

Comparison	Maximum P-value	Conclusion (alpha=5%)
A v B	0.3906	Noninferiority has not been demonstrated.
A v C	0.0457	Noninferiority has been demonstrated.
B v C	0.0040	Noninferiority has been demonstrated.

For each pair of standard deviations, the output displays an upper confidence bound for the ratio. If the upper confidence bound is greater than the upper equivalence differential, the P-value of an upper-tailed test comparing the ratio to Δ_u 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 ratio of each pair of standard deviations:

Noninferiority Test - alpha = 5%
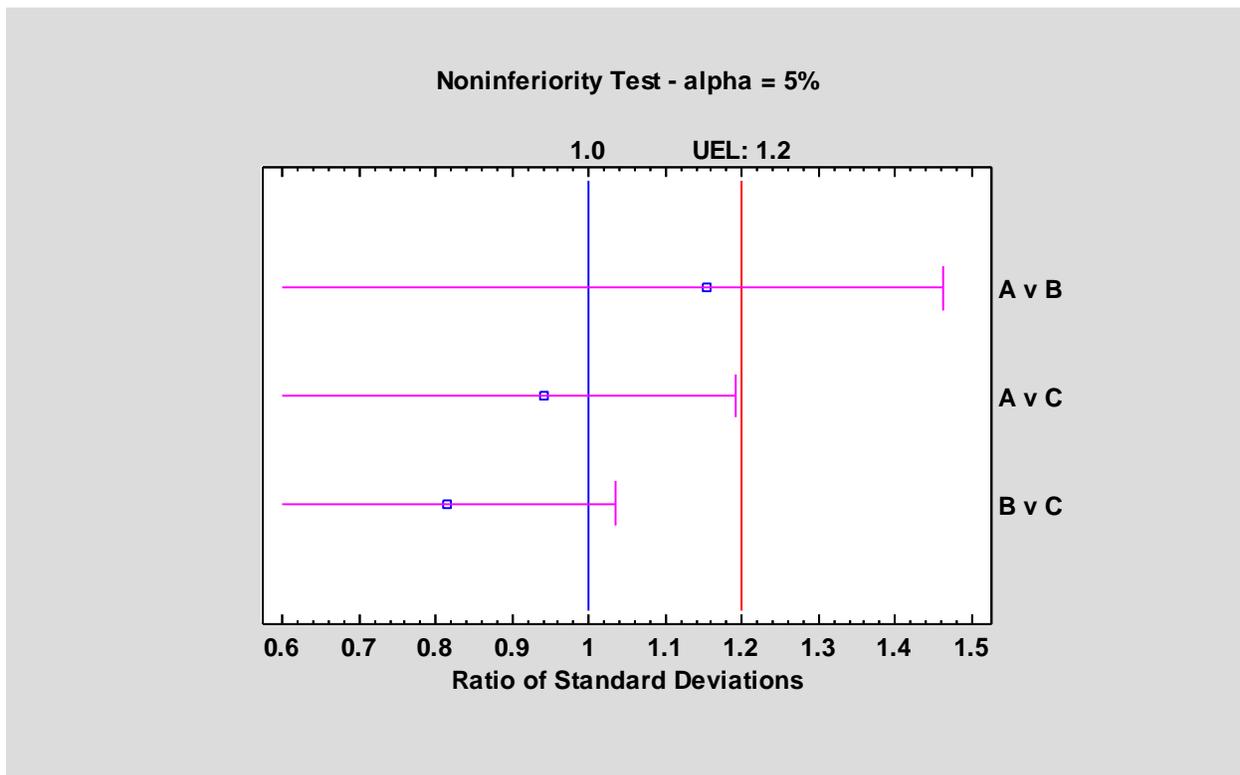
Noninferiority may be asserted for any ratios in which the confidence bounds do not contain the UEL.

Calculations

By default, the confidence intervals are calculated by:

$$\left[min\left(0, \frac{s_1^2}{s_2^2} \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}} \right), max\left(0, \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1} \right) \right]$$

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

$$\left[\frac{s_1^2}{s_2^2} \frac{1}{F_{\alpha/2, n_1-1, n_2-1}}, \frac{s_1^2}{s_2^2} F_{\alpha/2, n_2-1, n_1-1} \right]$$

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). <u>Statistics in Drug Research: Methodologies and Recent Developments</u>. 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.