Two Sample Comparison

Summary

The **Two Sample Comparison** procedure is designed to compare two independent samples of variable data. Tests are run to determine whether or not there are significant differences between the means, variances, and/or medians of the populations from which the samples were taken. In addition, the data may be displayed graphically in various ways, including a dual histogram, a multiple box-and-whisker plot, and a quantile-quantile plot.

In this procedure, it is assumed that no relationship exists between a selected observation in one sample and any specific observation in the other sample. In cases where specific observations in each sample correspond to the same experimental unit, use the *Paired Sample Comparison* procedure instead.

Sample StatFolio: *two samples.sgp*

Sample Data:

The file *bloodpressure.sgd* contains data describing the mean arterial blood pressure of 27 hypertension patients, $n_1 = 15$ of which were treated with a placebo and $n_2 = 12$ of which were given an experimental drug. The data have been entered into 2 columns as shown below:

Placebo	Test Agent
113	112
116	106
147	100
114	100
111	100
132	96
125	97
109	98
107	104
114	93
116	96
98	98
123	
126	
123	

Alternatively, all 27 blood pressure readings could have been entered into a single data column and a second column created identifying which group each patient belonged to:

| Patient	Pressure	Group
1	113	Placebo
2	116	Placebo
3	147	Placebo
4	114	Placebo
5	111	Placebo
6	132	Placebo
7	125	Placebo
8	109	Placebo
9	107	Placebo
10	114	Placebo
11	116	Placebo
12	98	Placebo
13	123	Placebo
14	126	Placebo
15	123	Placebo
16	112	Test Agent
17	106	Test Agent
18	100	Test Agent
19	100	Test Agent
20	100	Test Agent
21	96	Test Agent
22	97	Test Agent
23	98	Test Agent
24	104	Test Agent
25	93	Test Agent
26	96	Test Agent
27	98	Test Agent
28	112	Test Agent

Data Input

The data to be analyzed is specified on the dialog box shown below. For cases were the two samples have been placed in separate columns, specify the names of the two columns.
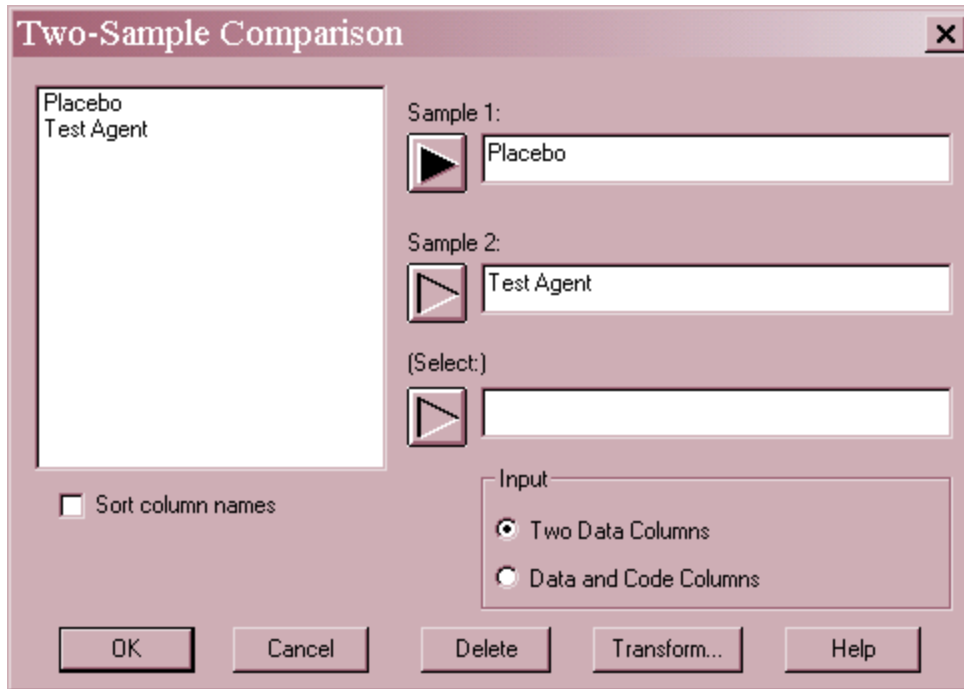
- **Sample 1:** numeric column containing the observations in the first sample.

- **Sample 2:** numeric column containing the observations in the second sample.

- **Select:** subset selection.

- **Input**: set to *Two Data Columns* to indicate that each sample has been placed in a separate column.

If the data from both samples have been entered into a single column, then enter the name of that column and the column containing the group identifiers:
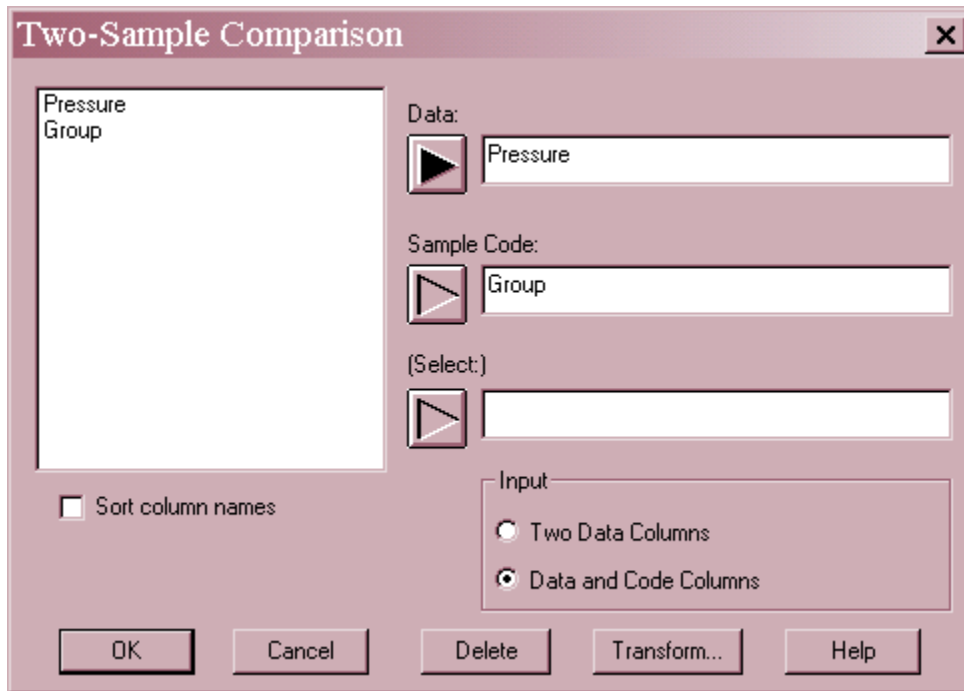
- **Data:** numeric column containing the observations from both samples.

- **Sample Code:** numeric or non-numeric column containing a identifier for the sample corresponding to each observation.

- **Select:** subset selection.

- **Input**: set to *Data and Code Columns* to indicate that the data for both samples have been placed into a single column.

Analysis Summary

The *Analysis Summary* shows the number of observations in each sample.

Two-Sample Comparison - Placebo & Test Agent
Sample 1: Placebo
Sample 2: Test Agent
Sample 1: 15 values ranging from 98.0 to 147.0
Sample 2: 12 values ranging from 93.0 to 112.0

Also displayed are the largest and smallest values.

Summary Statistics

The *Summary Statistics* pane calculates a number of different statistics that are commonly used to summarize a sample of variable data:

Summary Statistics		
	Placebo	Test Agent
Count	15	12
Average	118.267	100.0
Median	116.0	99.0
Mode		100.0
Geometric mean	117.745	99.8812
5% Trimmed mean	117.796	99.7222
5% Winsorized mean	118.267	100.0
Variance	136.781	26.7273
Standard deviation	11.6953	5.16984
Coeff. of variation	9.88896%	5.16984%
Standard error	3.01972	1.49241
5% Winsorized sigma	11.6953	5.16984
MAD	7.0	2.5
Sbi	10.7038	4.52375
Minimum	98.0	93.0
Maximum	147.0	112.0
Range	49.0	19.0
Lower quartile	111.0	96.5
Upper quartile	125.0	102.0
Interquartile range	14.0	5.5
1/6 sextile	109.0	96.0
5/6 sextile	126.0	105.0
Intersextile range	17.0	9.0
Skewness	0.82607	1.17953
Stnd. skewness	1.30613	1.6681
Kurtosis	1.63626	1.55061
Stnd. kurtosis	1.29358	1.09645
Sum	1774.0	1200.0
Sum of squares	211720.0	120294.0

Most of the statistics fall into one of three categories:

1. measures of *central tendency* – statistics that characterize the "center" of the data.
2. measure of *dispersion* – statistics that measure the spread of the data.
3. measures of *shape* – statistics that measure the shape of the data relative to a normal distribution.

The statistics included in the table by default are controlled by the settings on the *Stats* pane of the *Preferences* dialog box. Within the procedure, the selection may be changed using *Pane Options*. For a detailed description of each statistic, see the *One Variable Analysis* documentation.

For the data on blood pressure, note the following sample means and standard deviations:

Placebo: $\bar{x}_1 = 118.3$ $s_1 = 11.70$

Test Agent: $\bar{x}_2 = 100.0$ $s_2 = 5.17$

Both the mean and standard deviation appear to be somewhat higher for the sample that was given the placebo. Note also that the standardized skewness and standardized kurtosis for both samples are within the interval –2 to +2, indicating that both samples could reasonably be assumed to have come from normal distributions.
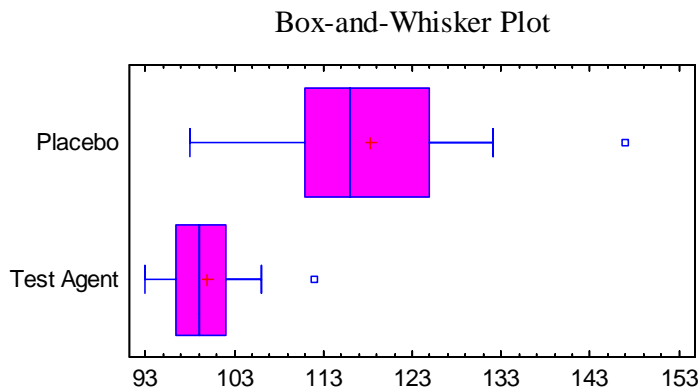
Pane Options
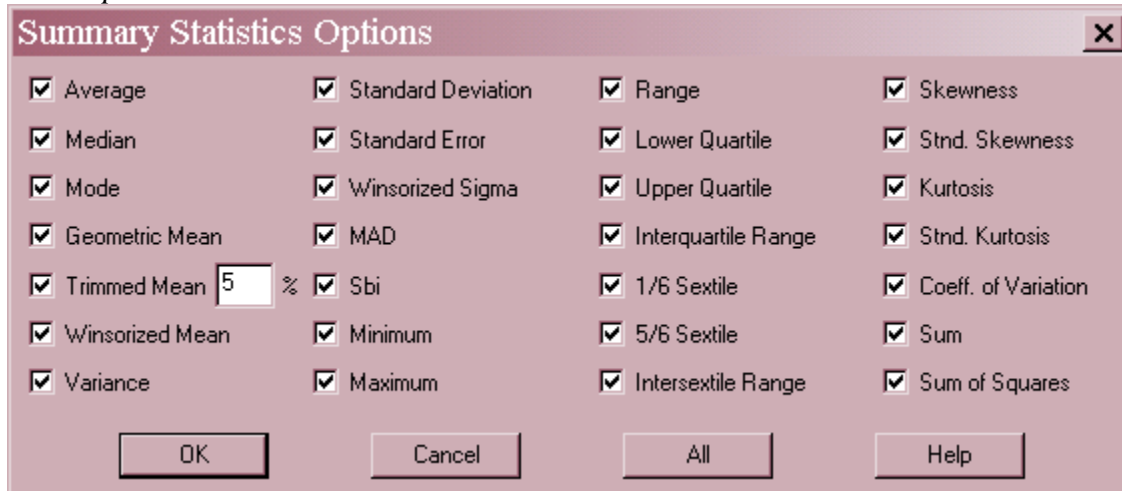
Select the desired statistics.

Box-and-Whisker Plot

This pane displays a box-and-whisker plot for each sample.

Box-and-whisker plots are constructed in the following manner:

- A box is drawn extending from the *lower quartile* of the sample to the *upper quartile*. This is the interval covered by the middle 50% of the data values when sorted from smallest to largest.

- A vertical line is drawn at the *median* (the middle value).

- If requested, a plus sign is placed at the location of the sample mean.

- Whiskers are drawn from the edges of the box to the largest and smallest data values, unless there are values unusually far away from the box (which Tukey calls *outside points*). Outside points, which are points more than 1.5 times the interquartile range (box width) above or below the box, are indicated by point symbols. Any points more than 3 times the interquartile range above or below the box are called *far outside points*, and are indicated by point symbols with plus signs superimposed on top of them. If outside points are present, the whiskers are drawn to the largest and smallest data values which are not outside points.

In the sample data, the box for the Placebo group is shifted to the right of the box for the group given the Test Agent. In addition, the box is wider, corresponding to the greater variability. Each of the plots also shows a single outside point.

Pane Options
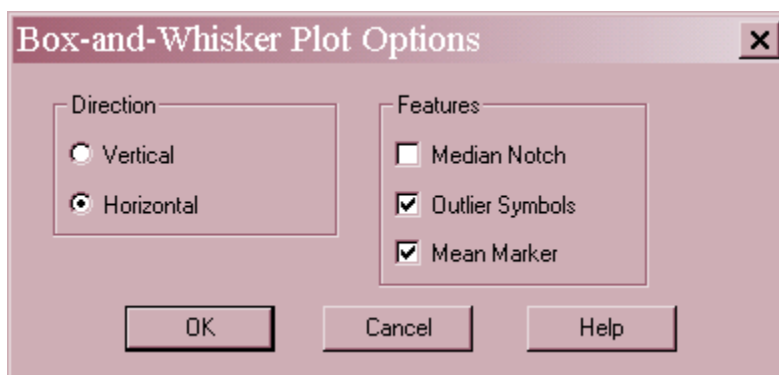
- **Direction**: the orientation of the plot, corresponding to the direction of the whiskers.

- **Median Notch**: If selected, a notch will be added to the plot showing the estimation error associated with each median. The notches are scaled in such a way that if they do not overlap, the two medians are significantly different at the default system confidence level (set on the *General* tab of the *Preferences* dialog box on the *Edit* menu).

- **Outlier Symbols**: if selected, indicates the location of outside points.

- **Mean Marker**: if selected, shows the location of the sample mean as well as the median.

Example – Notched Box-and-Whisker Plot

The following plot adds median notches at the 95% confidence level.

Box-and-Whisker Plot
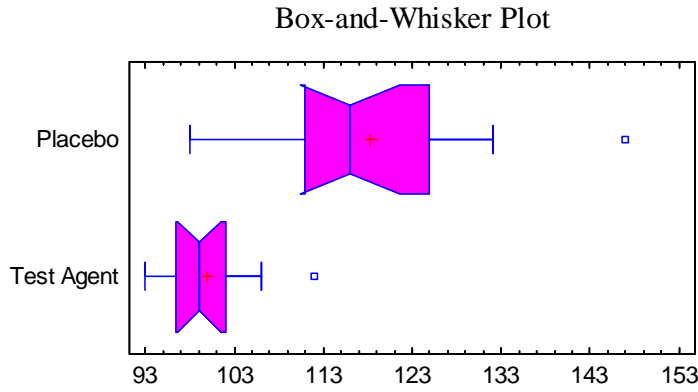
Each notch covers the interval

$$\tilde{x}_j \pm \frac{z_{\alpha/2}}{2} \frac{1.25(IQR_j)}{1.35\sqrt{n_j}} \left(1 + \frac{1}{\sqrt{2}}\right) \tag{1}$$

where $\tilde{x}_j$ is the median of the *j*-th sample, IQR_j is the sample interquartile range, n_j is the sample size, and $z_{\alpha/2}$ is the upper $(\alpha/2)\%$ critical value of a standard normal distribution. Since the notches do not overlap, the medians are significantly different at the 5% significance level.

Frequency Histogram

The *Frequency Histogram* pane displays the sample data using a dual barchart:
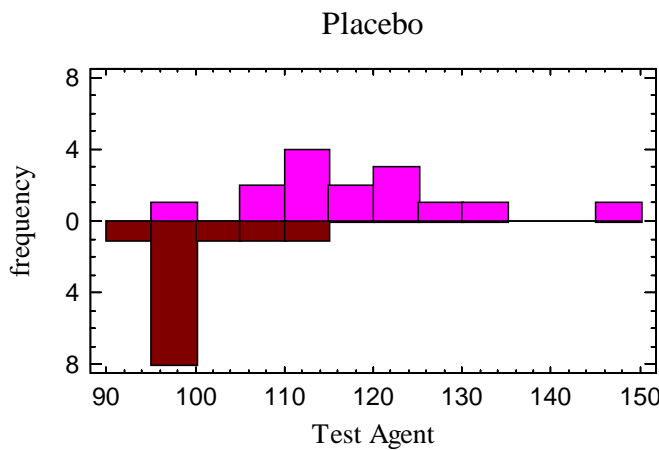
Placebo

The height of each bar in the plot above represents the number of observations in adjacent intervals, each covering 5 mm of pressure. The histogram above the line is for the group given the placebo, while the histogram below the line is for the group given the test agent.
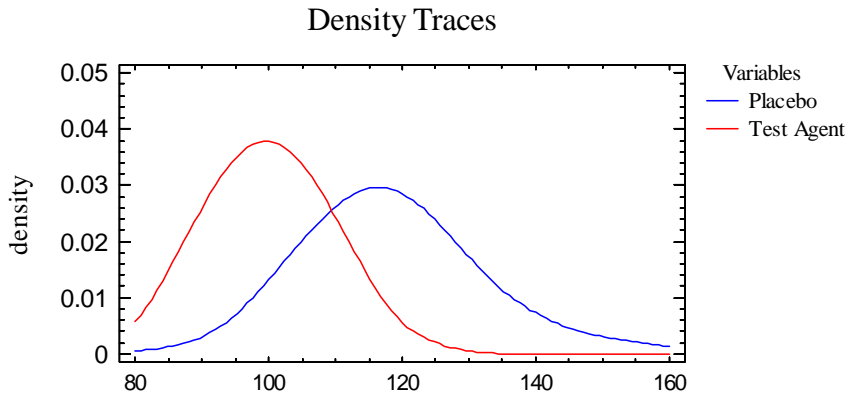
Pane Options

- **Number of classes**: the number of intervals into which the data will be divided. Intervals are adjacent to each other and of equal width.

- **Lower Limit**: lower limit of the first interval.

- **Upper Limit**: upper limit of the last interval.

- **Hold**: maintains the selected number of intervals and limits even if the source data changes. By default, the number of classes and the limits are recalculated whenever the data changes. This is necessary so that all observations are displayed even if some of the updated data fall beyond the original limits.

- **Counts**: if *Relative*, the height of the bars represents the observations in a single interval. If *Cumulative*, the height represents the observations in the indicated interval and all intervals to its left.

Density Trace

The *Density Trace* provides a nonparametric estimate of the probability density function of the populations from which the data were sampled. It is created by counting the number of observations which fall within a window of fixed width moved across the range of the data.

Density Traces

The estimated density function is given by:

$$f(x) = \frac{1}{hn} \sum_{i=1}^{n} W\left(\frac{x - x_i}{h}\right) \tag{2}$$

where *h* is the width of the window in units of *X* and *W(u)* is a weighting function determined by the selection on the *Pane Options* dialog box. Two forms of weighting function are offered:
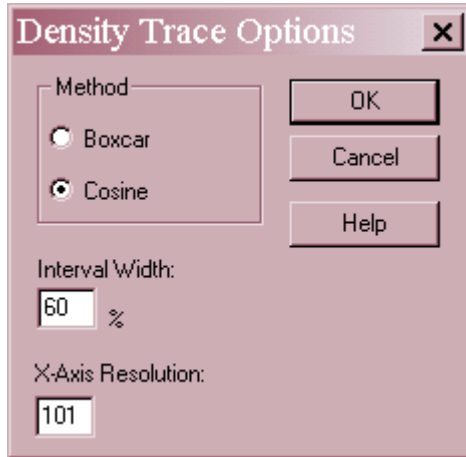
Boxcar Function

$$W(u) = \begin{cases} 1 & if \, |u| \leq 1/2 \\ 0 & otherwise \end{cases} \tag{3}$$

Cosine Function

$$W(u) = \begin{cases} 1 + \cos(2\pi u) & if \, |u| < 1/2 \\ 0 & otherwise \end{cases} \tag{4}$$

The latter selection usually gives a smoother result, with the desirable value of *h* depending on the size of the data sample.

For the sample data, the density traces closely resemble normal distributions.

Pane Options

- **Method:** the desired weighting function. The boxcar function weights all values within the window equally. The cosine function gives decreasing weight to observations further from the center of the window. The default selection is determined by the setting on the *EDA* tab of the *Preferences* dialog box accessible from the *Edit* menu.

- **Interval Width:** the width of the window *h* within which observations affect the estimated density, as a percentage of the range covered by the x-axis. *h* = 60% is not unreasonable for a small sample but may not give as much detail as a smaller value in larger samples.

- **X-Axis Resolution**: the number of points at which the density is estimated.

Comparison of Standard Deviations

The plots displayed above suggest that differences exist between the two samples. To determine whether the apparent differences are statistically significant and not just due to the small sample sizes, hypothesis tests may be performed. The *Comparison of Standard Deviations* performs an F test to determine whether or not the standard deviations are significantly different by considering the variance ratio defined by:

$$\varpi = \frac{\sigma_1^2}{\sigma_2^2} \tag{5}$$

The output is shown below:

Comparison of Standard Deviations

	Placebo	Test Agent
Standard deviation	11.6953	5.16984
Variance	136.781	26.7273
Df	14	11

Ratio of Variances = 5.11765

95.0% Confidence Intervals
   Standard deviation of Placebo: [8.56247,18.4447]
   Standard deviation of Test Agent: [3.66229,8.77776]
   Ratio of Variances: [1.52365,15.837]

F-test to Compare Standard Deviations
  Null hypothesis: sigma1 = sigma2
  Alt. hypothesis: sigma1 NE sigma2
  F = 5.11765 P-value = 0.00988302
  Reject the null hypothesis for alpha = 0.05.

The table shows:

1. *Sample Statistics:* the sample standard deviations, variances, and degrees of freedom.

2. *Ratio of Variances:* the ratio of the variance of the first sample to the variance of the second sample, given by:

$$\hat{\omega} = \frac{s_1^2}{s_2^2} \tag{6}$$

For the blood pressure data, the variance in the first sample is over 5 times the variance in the second sample.

3. *Confidence Intervals:* interval estimates for each sample variance and for the ratio ω. The interval for the ratio indicates that it could actually be anywhere between approximately 1.5 and 15.8, with 95% confidence. Note that the small sample sizes do not provide much precision in estimating that variance ratio.

4. *F-Test*: a test of hypotheses concerning the variance ratio. By default, the hypotheses tested are:

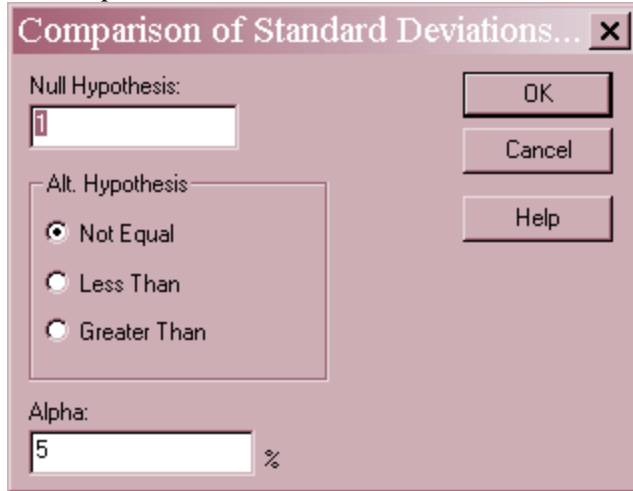   *Null Hypothesis*: $\omega = 1$

      Alt. Hypothesis: $\omega \neq 1$

although this can be changed using *Pane Options*. To test the hypothesis, it computes the F-ratio:

$$F = \frac{s_1^2 / s_2^2}{\varpi_0} \tag{7}$$

where ω_0 is the value specified by the null hypothesis. Small P-values (less than 0.05 if operating at the 5% significance level) lead to a rejection of the null hypothesis. In the current example, the small P-Value indicates that the samples come from populations with significantly different standard deviations.

Pane Options

- **Null Hypothesis:** ω_0, the value of the variance ratio specified by the null hypothesis.

- **Alt**. **Hypothesis:** the alternative hypothesis may be two-sided ("Not Equal") or one-sided (such as ω > 1 if "Greater Than" is specified.).

- **Alpha:** the significance level of the test, usually set to 0.01, 0.05, or 0.10. This equals the probability of rejecting the null hypothesis if it is true. It does not affect the P-Value, only the conclusion stated immediately below the P-Value.

Comparison of Means

The *Comparison of Means* performs a t test to determine whether or not the means of the two samples are significantly different by considering the difference defined by:

$$\Delta = \mu_1 - \mu_2 \tag{8}$$

The output is shown below:

Comparison of Means
95.0% confidence interval for mean of Placebo: 118.267 +/- 6.47668 [111.79,124.743]
95.0% confidence interval for mean of Test Agent: 100.0 +/- 3.28477 [96.7152,103.285]
95.0% confidence interval for the difference between the means
  not assuming equal variances: 18.2667 +/- 7.0231 [11.2436,25.2898]

t test to compare means
  Null hypothesis: mean1 = mean2
  Alt. hypothesis: mean1 NE mean2
    not assuming equal variances: t = 5.42298   P-value = 0.0000255718
  Reject the null hypothesis for alpha = 0.05.

The table displays:

1. *Confidence Intervals:* interval estimates for each sample mean and for the difference between the means Δ. The interval for the difference indicates that the mean blood pressure of the Placebo group could exceed that of the test agent group by anywhere between approximately 11.2 and 25.3, with 95% confidence.

2. *t-Test*: a test of hypotheses concerning the difference between the means. By default, the hypotheses tested are:
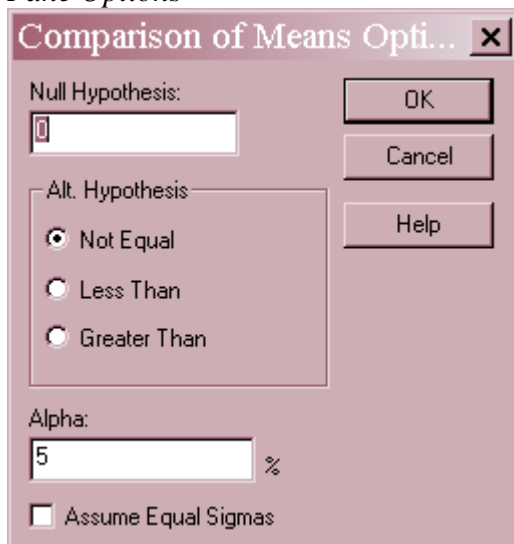
   Null Hypothesis: $\Delta = 0$

   Alt. Hypothesis: $\Delta \neq 0$

   although this can be changed using *Pane Options*. To test the hypothesis, it computes a t-statistic. Depending upon *Pane Options*, an exact t-test may be run assuming that the samples come from populations with equal variances, or an approximate t-test may be run without making that assumption. Associated with each t-statistic is a P-Value. Small P-values (less than 0.05 if operating at the 5% significance level) lead to a rejection of the null hypothesis.

Since the F-test described earlier indicated that there was a significant difference between the standard deviations of the two samples, the t-test was run without assuming equal variances. The extremely small P-value indicates that the samples come from populations with significantly different means.

Pane Options

- **Null Hypothesis:** Δ_0, the value of the difference between the means specified by the null hypothesis.

- **Alt**. **Hypothesis:** the alternative hypothesis may be two-sided ("Not Equal") or one-sided (such as $\Delta > 1$ if "Greater Than" is specified.).

- **Alpha:** the significance level of the test, usually set to 0.01, 0.05, or 0.10. This equals the probability of rejecting the null hypothesis if it is true. It does not affect the P-Value, only the conclusion stated immediately below the P-Value.

- **Assume Equal Sigmas**: whether or not to assume that the two samples come from populations with equal variances. The t-test is exact if the assumption is made and approximate if it is not.

Comparison of Medians

STATGRAPHICS also provides a tabular option that compares the medians of two populations using the Mann-Whitney (Wilcoxon) test. This test is useful in cases where the assumption of normality is questionable and no suitable transformation of the data can be found, and in cases where one or more observations are suspected to be outliers. To perform the test, the two samples are combined and ranked from smallest to largest (from rank 1 to rank n_1+n_2), with any tied observations being given the average rank for the values in the tied group.   The output displayed is shown below:

```
Comparison of Medians
Median of sample 1: 116.0
Median of sample 2: 99.0

Mann-Whitney (Wilcoxon) W test to compare medians
   Null hypothesis: median1 = median2
   Alt. hypothesis: median1 NE median2

   Average rank of sample 1: 19.3333
   Average rank of sample 2: 7.33333

   W = -80.0   P-value = 0.000101812
Reject the null hypothesis for alpha = 0.05.
```

The table shows:

1. *Sample Statistics*: the medians of the two samples.

2. *Average Ranks*: the average rank of the data in each sample when both samples are ranked together.

3. *W-Test:* the Wilcoxon test statistic and its associated P-Value. Small P-values (less than 0.05 if operating at the 5% significance level) lead to a rejection of the null hypothesis.

The small P-Value in the current example indicates that the samples come from populations with significantly different medians.

Pane Options
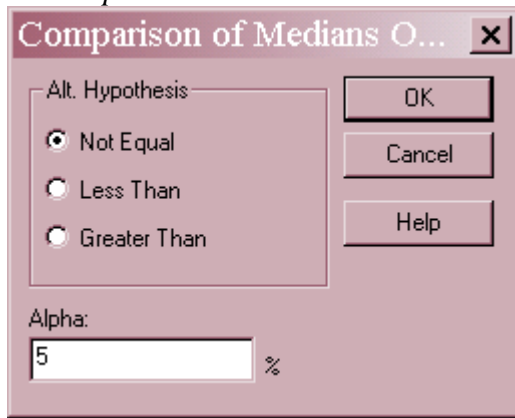
- **Alt**. **Hypothesis:** the alternative hypothesis may be two-sided ("Not Equal") or one-sided (such as *median₁ > median₂* if "Greater Than" is specified.).

- **Alpha:** the significance level of the test, usually set to 0.01, 0.05, or 0.10. This equals the probability of rejecting the null hypothesis if it is true. It does not affect the P-Value, only the conclusion stated immediately below the P-Value.

Quantile Plot

This pane plots the quantiles (percentiles) of the data in each sample.
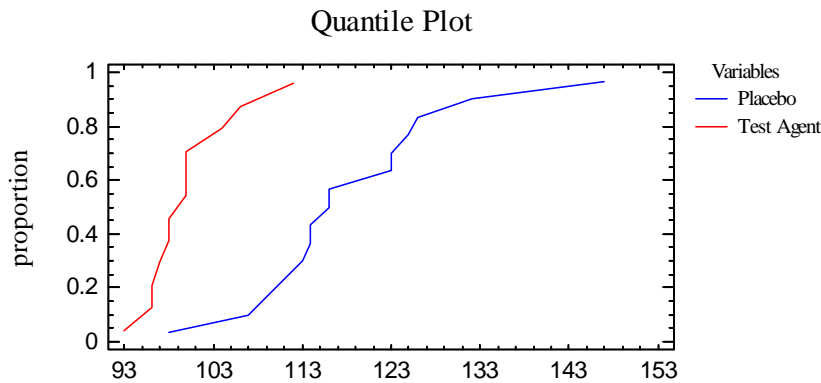
In this plot, the data are sorted from smallest to largest and plotted at the coordinates

$$\left(x_{(i)}, \frac{i - 0.5}{n_j} \right) \tag{9}$$

The offset between the two lines corresponds to the different locations of the two distributions.

Quantile-Quantile Plot

The quantile-quantile (Q-Q) plot plots the quantiles of one sample versus the quantiles of the other.

Quantile-Quantile Plot
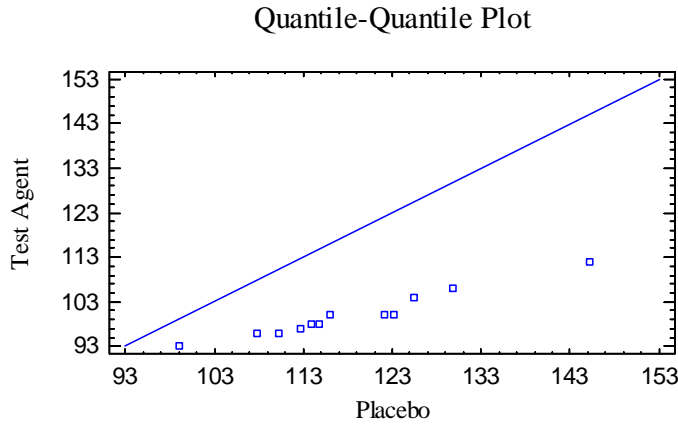
In the plot, points are drawn for each observation in the smaller sample versus interpolated percentiles for the larger sample. If the two samples come from the same underlying population, the points should lie approximately along the diagonal line. The offset of the points to the right of the line corresponds to the larger mean of the Placebo group, while the fact that the slope appears to be less than 1 is an indication that the Placebo group also has a larger standard deviation.

Kolmogorov-Smirnov Test

The *Kolmogorov-Smirnov Test* performs a formal test of the null hypothesis that the two samples come from the same population. It does so by calculating the maximum distance between the empirical distributions

$$D_n = \max_x \left| F_1(x) - F_2(x) \right| \tag{10}$$

which equates to the maximum distance between the two curves in the *Quantile Plots*. The output is shown below:

Kolmogorov-Smirnov Test
Estimated overall statistic DN = 0.85
Two-sided large sample K-S statistic = 2.19469
Approximate P value = 0.000131017

In addition to D_n, the table also shows the K-S statistic defined by

$$K = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_n \tag{11}$$

and an associated P-Value. Small P-values (less than 0.05 if operating at the 5% significance level) lead to a rejection of the hypothesis that the two samples come from the same population.

<u>Calculations</u>

Confidence Interval for the Difference Between the Means

If equal variances are assumed:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{12}$$

where

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} \tag{13}$$

If equal variances are not assumed:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,m} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{14}$$

where

$$\frac{1}{m} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \tag{15}$$

and

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2} \tag{16}$$

Confidence Interval for the Variance Ratio

$$\left[\frac{s_1^2}{s_2^2} \frac{1}{F_{\alpha/2,n_1-1,n_2-1}}, \frac{s_1^2}{s_2^2} F_{\alpha/2,n_2-1,n_1-1} \right] \tag{17}$$

t-Test

If equal variances are assumed:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{18}$$

is compared to a t distribution with n_1+n_2-2 degrees of freedom. If equal variances are not assumed,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{19}$$

is compared to a t-distribution with m degrees of freedom.

Mann-Whitney (Wilcoxon) Test

$$W = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_x \tag{20}$$

where T_x is the sum of the ranks for sample 1. A normalized test statistic is computed from

$$Z = \frac{W - \frac{n_1 n_2}{2} - 0.5}{\text{var}(W)} \qquad \text{if } W > \frac{n_1 n_2}{2} \tag{21}$$

or

$$Z = \frac{W - \frac{n_1 n_2}{2} + 0.5}{\text{var}(W)} \qquad \text{if } W \leq \frac{n_1 n_2}{2} \tag{22}$$

In the above, the variance of W accounts for tied observations through

$$\text{var}(W) = \frac{n_1 n_2}{12} \left[n_1 + n_2 + 1 - \frac{\sum_{j=1}^{g} t_j (t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} \right] \tag{23}$$

where g is the number of tied groups and t_j is the size of tied group j. A P-value is computed by comparing Z to a standard normal distribution.

Kolmogorov-Smirnov P-Value

$$P = 1 \quad \text{if } K < 0.22 \tag{24}$$

$$P = 1 - \frac{\sqrt{2\pi}}{K} \exp\left(\frac{-\pi^2}{8K^2}\right) \quad \text{if } 0.22 < K \leq 0.80 \tag{25}$$

$$P = 2e^{-2K^2} + e^{-8K^2} - e^{-18K^2} \quad \text{if } 0.80 < K \leq 3.15 \tag{26}$$

$$P = 0 \quad \text{if } K > 3.15 \tag{27}$$