statgraphics

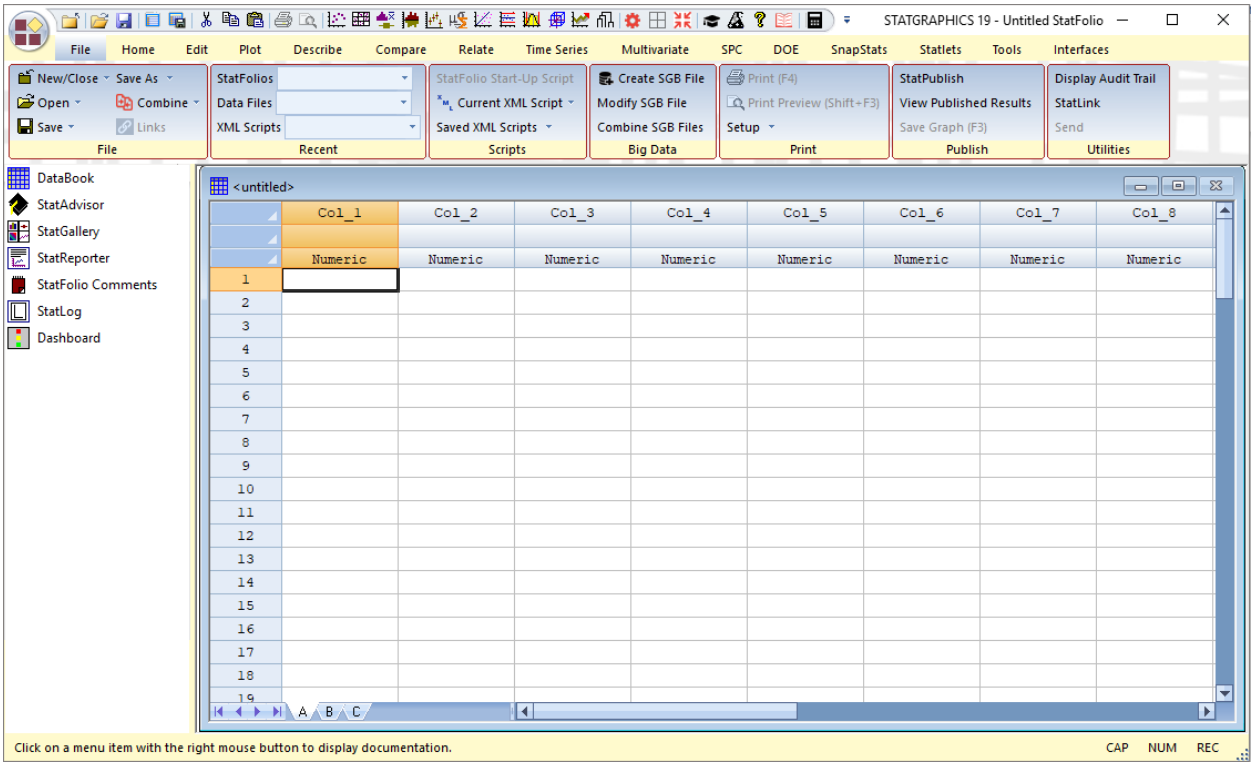
STATGRAPHICS Centurion Version 19 Enhancements

Statgraphics 19 contains many additions and enhancements. Significant changes have been made to the program, including a new user interface, 16 new statistical procedures, improvements to 40 existing statistical procedures, and an interface to Python. Many of these changes are illustrated below.

User Interface

Menu and Main Toolbar

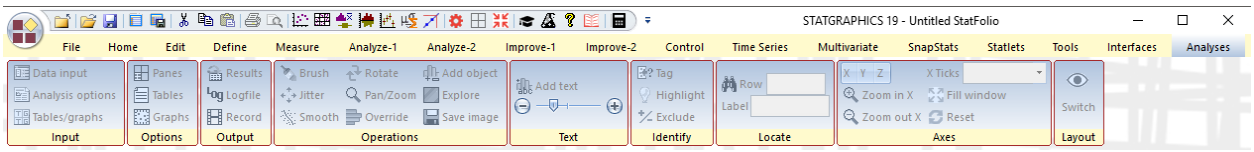
One of the most noticeable changes in Version 19 is a modernization of the user interface.



Changes include replacement of the old pulldown menus with a ribbon bar, which consists of 16 categories when using the standard menu and 17 categories when using the Six Sigma menu. Each category contains multiple panels that give access to both system options and statistical procedures. The Quick Access Toolbar has been moved to the top of the window for easy access to commonly used procedures.

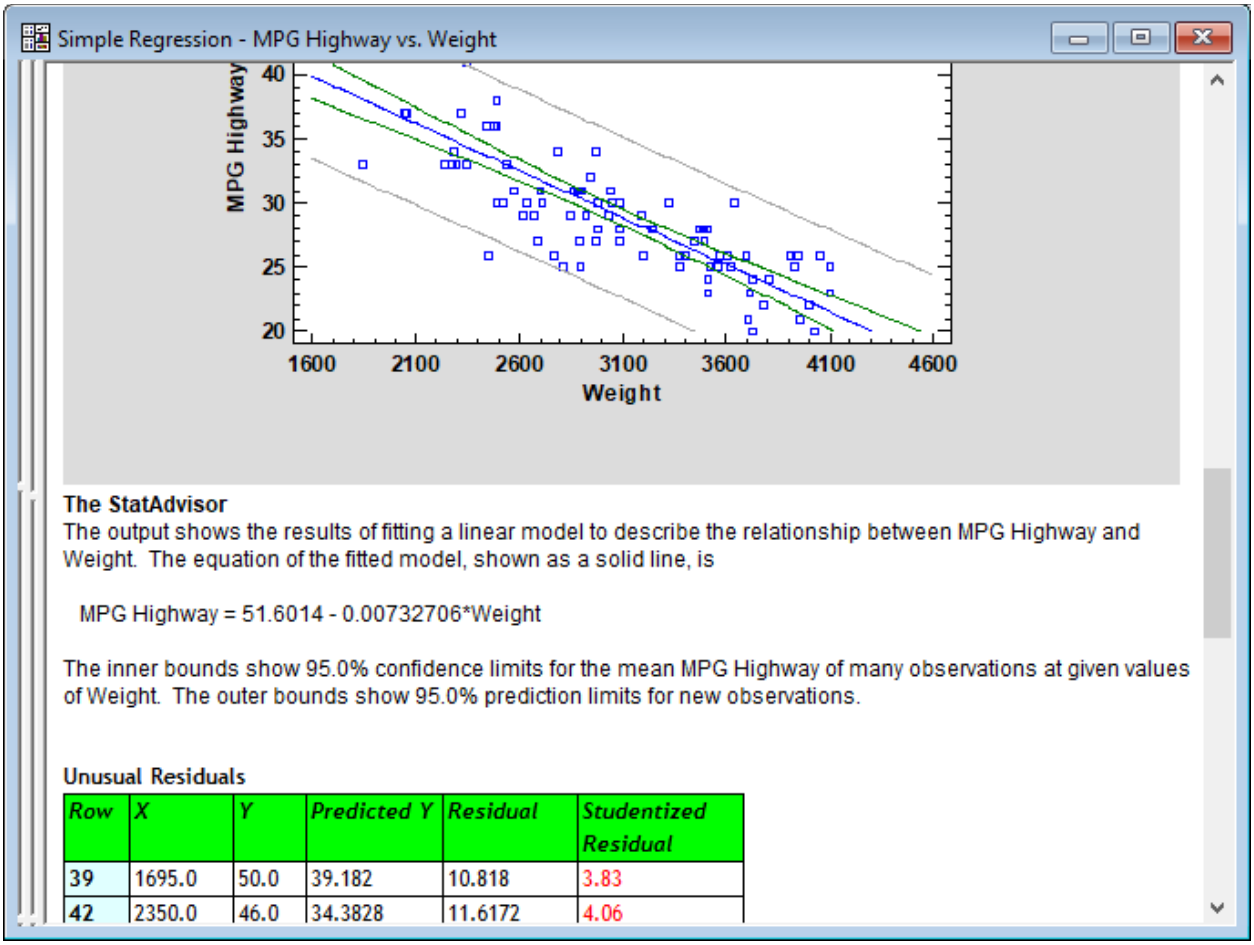
Analysis Toolbar

After an analysis has been performed, the ribbon bar displays panels that replace the old analysis toolbar. This leaves more real estate for displaying analysis windows.



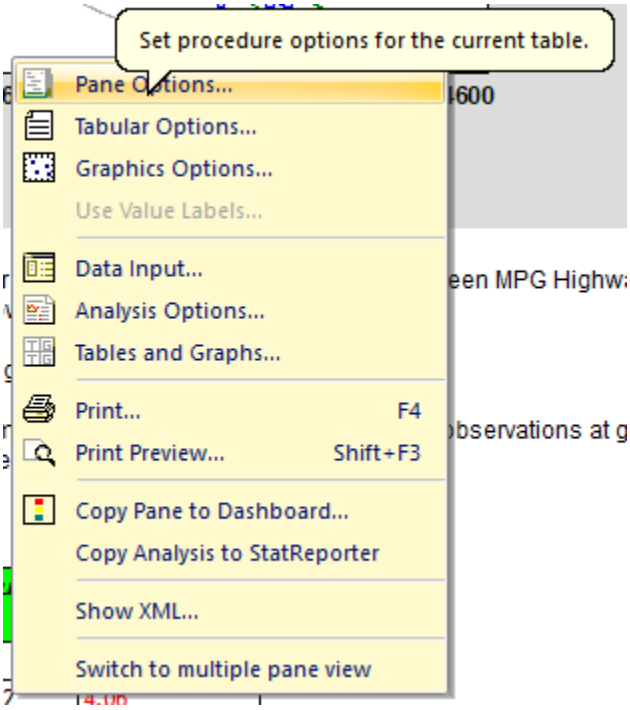
Single Pane View in Analysis Windows

Analysis windows now offer two layouts for displaying tables and graphs: the multiple pane view used in earlier versions and a single pane view displaying all tables and graphs in a report-style format. The *Switch* button on the *Layout* pane switches between the 2 layouts.



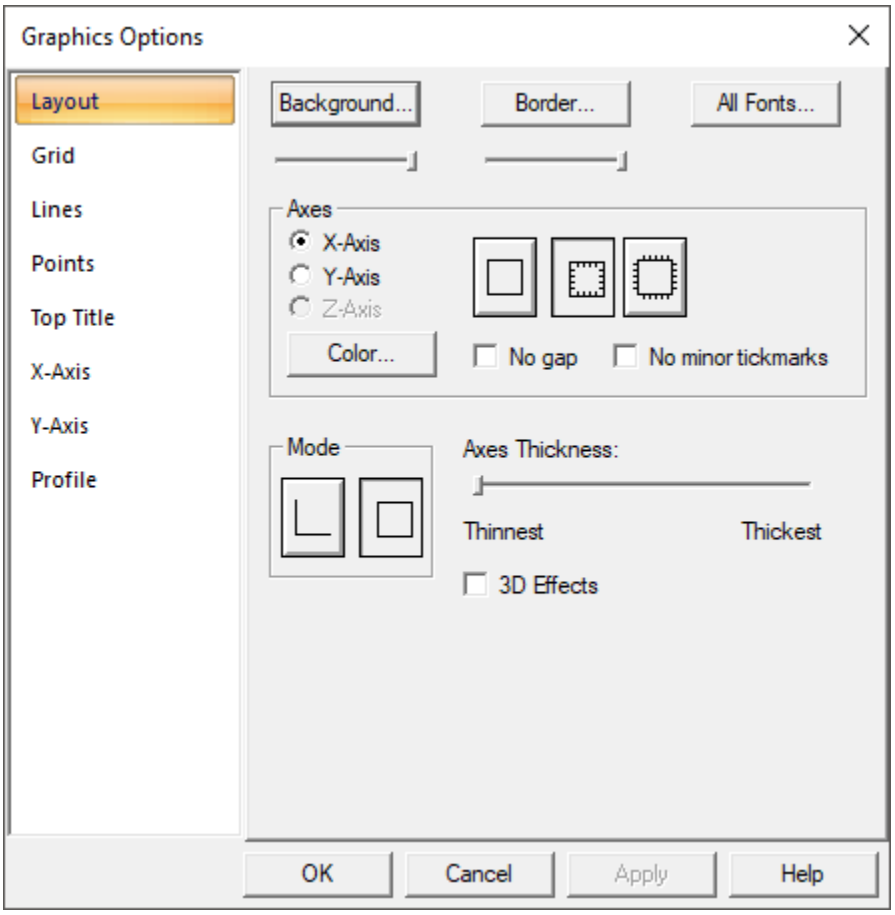
Popup Menus with Tooltips

Revised popup menus now display icons and tooltips.



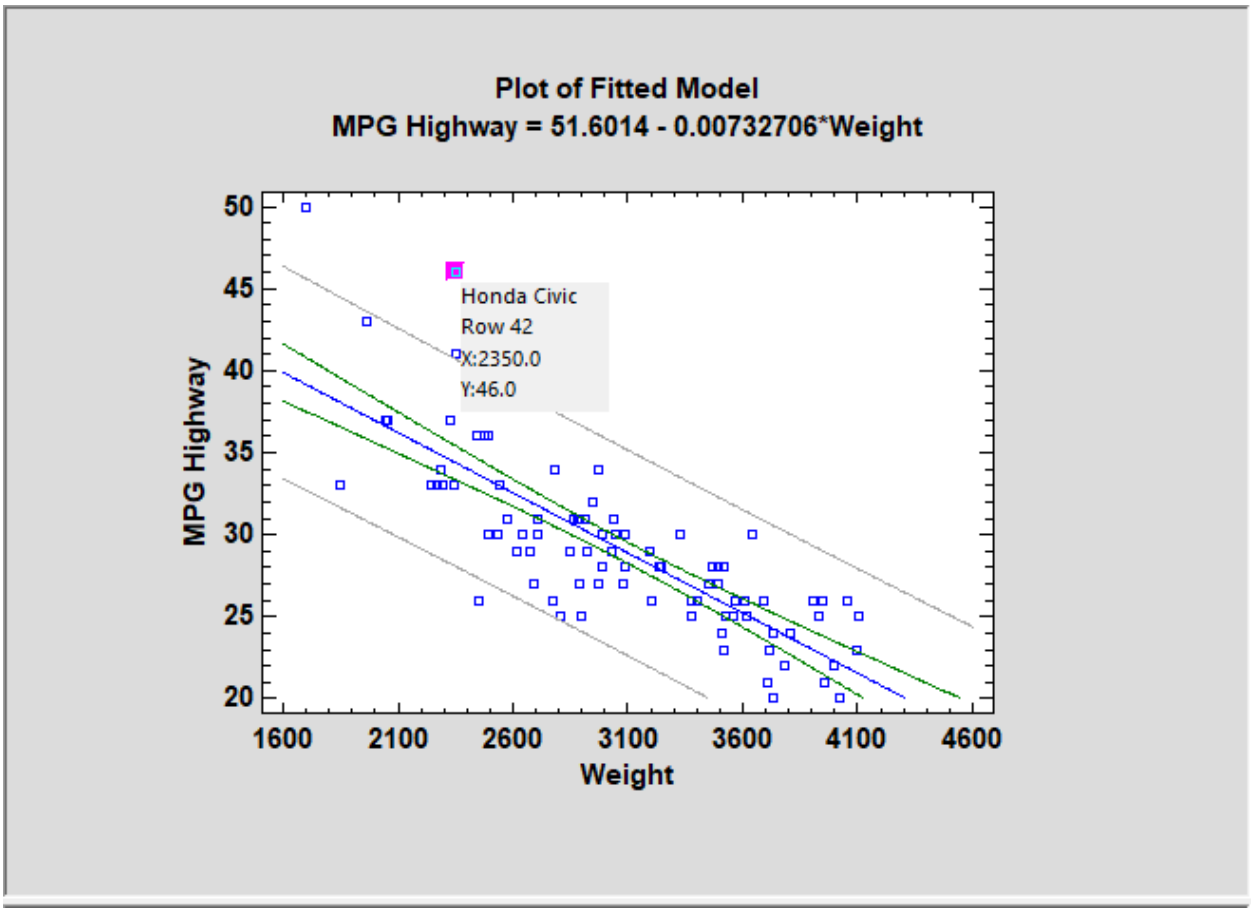
New Multi-Item Dialog Boxes

The tabbed dialog boxes used to specify system preferences, graphics options and StatFolio alerts have been replaced by dialog boxes containing a navigation bar along the left side.



Point Identification

When the mouse is held down on a point in a graph, the coordinates of that point are displayed together with other information.



System Dashboard

A new window called the *Dashboard* has been added to all StatFolios to display selected tables and graphs from multiple analysis windows. Each table and graph is linked to an analysis window and updated immediately whenever that analysis changes. Selected procedures such as control charts and regression models may be set to color the background of each pane in the dashboard using green, yellow or red to indicate alerts. For example, a red background on a control chart might indicate one or more points beyond the 3-sigma control limits.

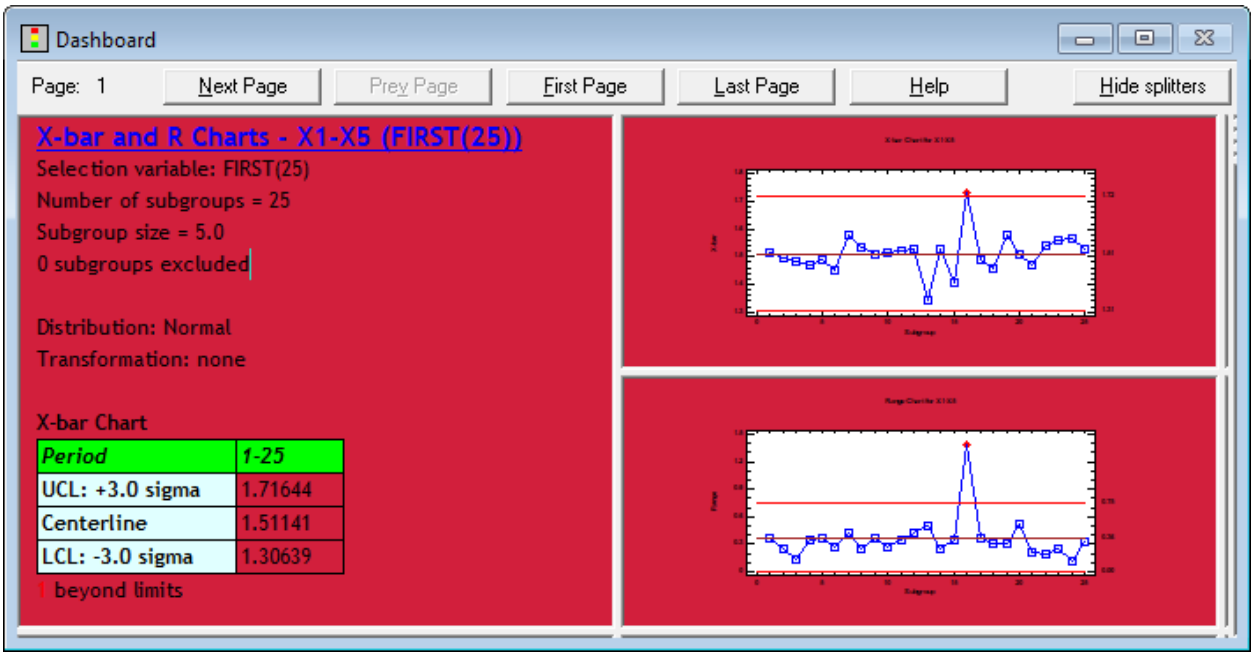
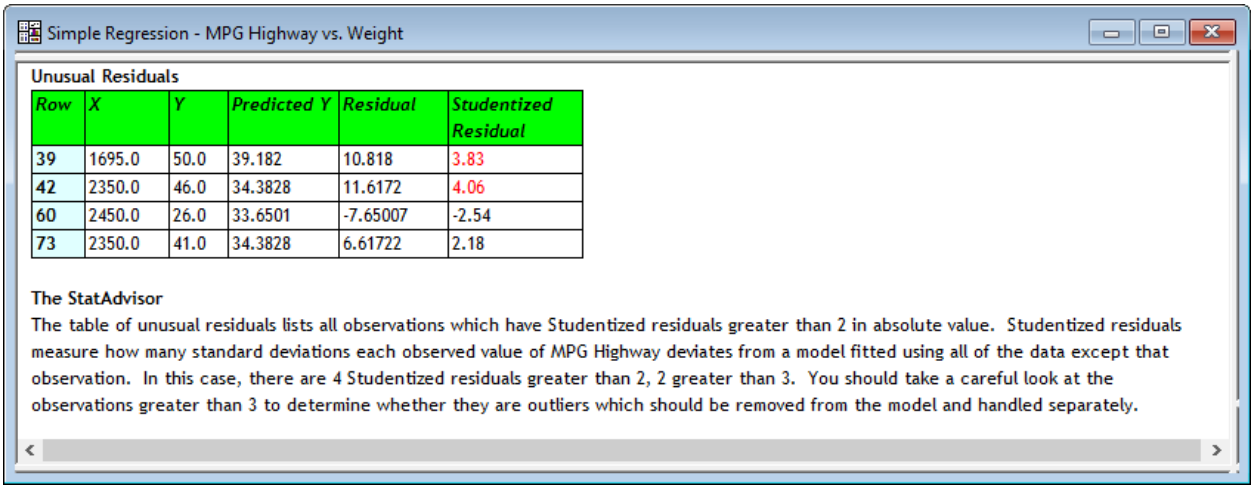
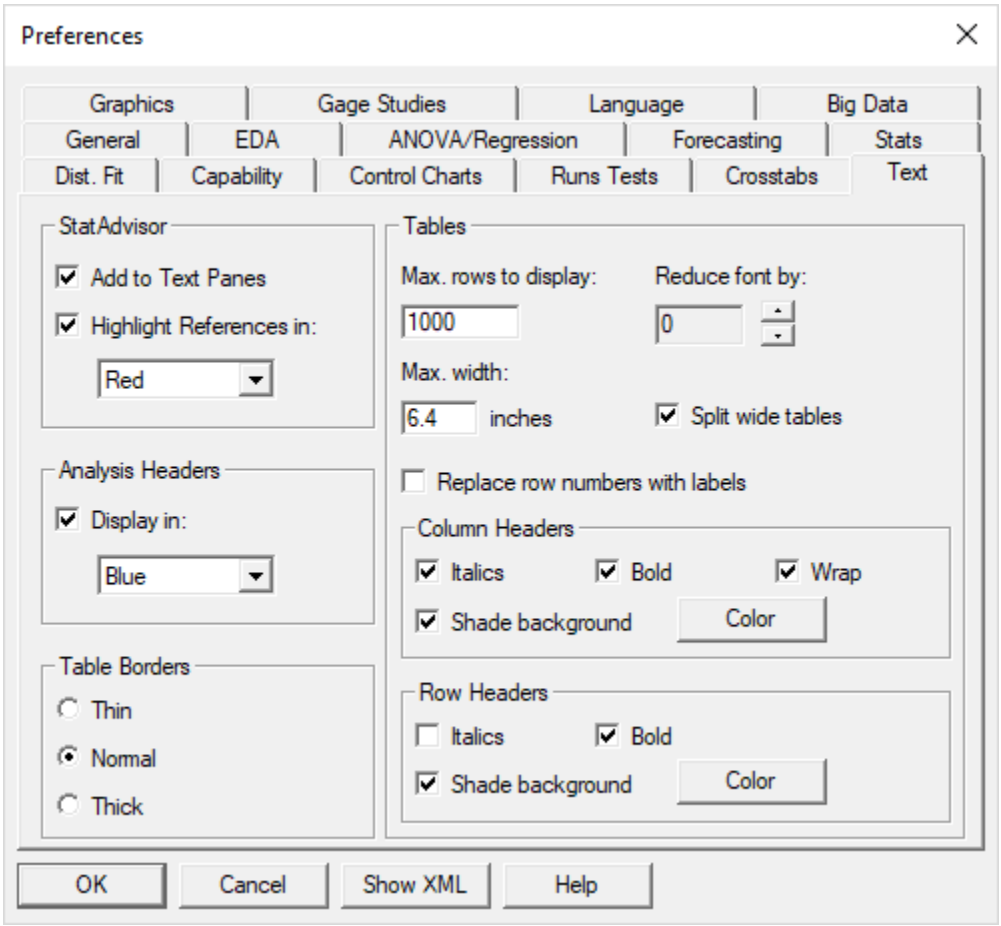


Table Row and Column Headers

Users may elect to highlight rows and/or columns in output tables.



The thickness of table borders may also be changed.

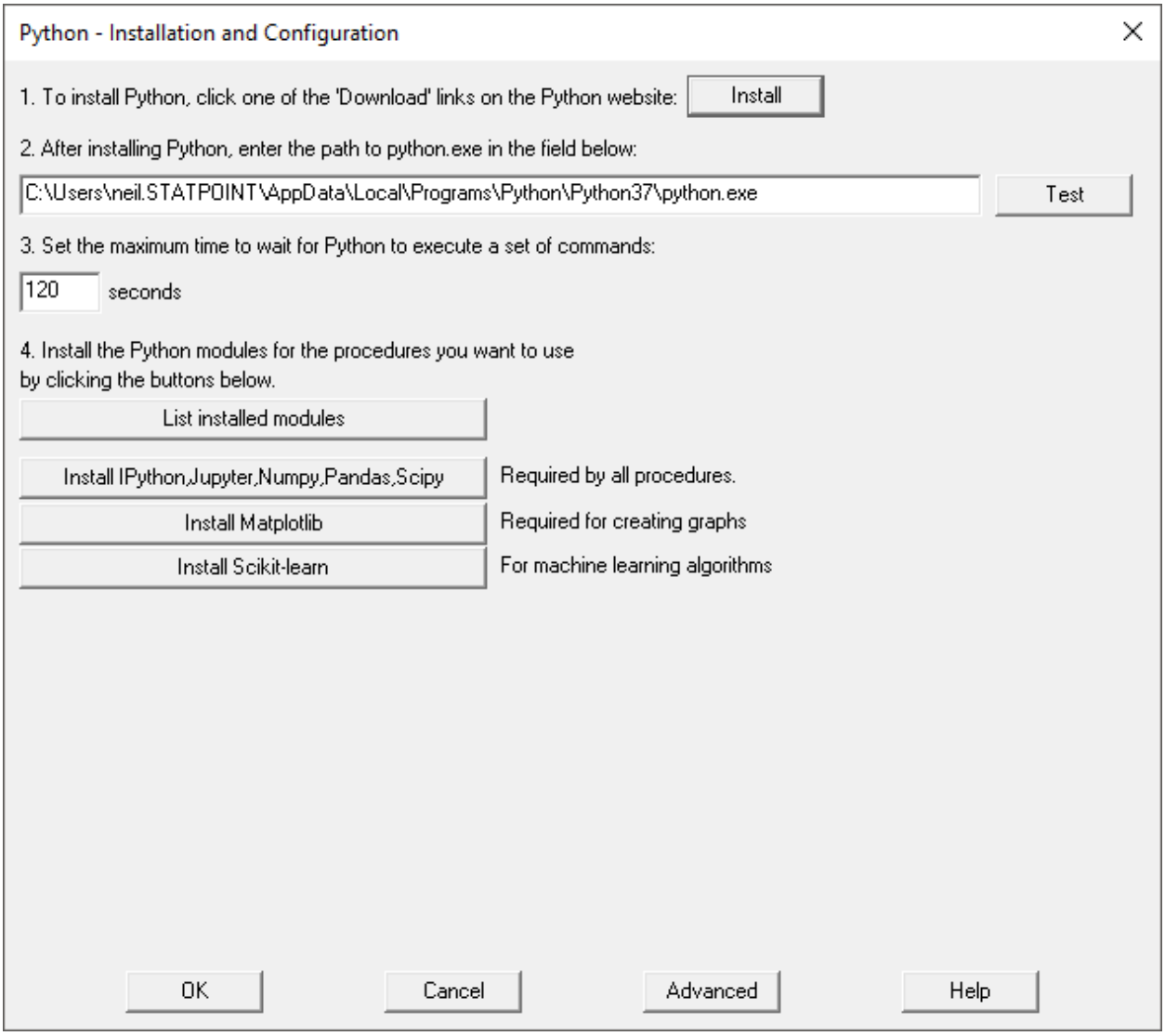


Python Interface

Statgraphics 19 includes an interface to the Python programming language. It is similar to the interface to R that was included in Statgraphics 18.

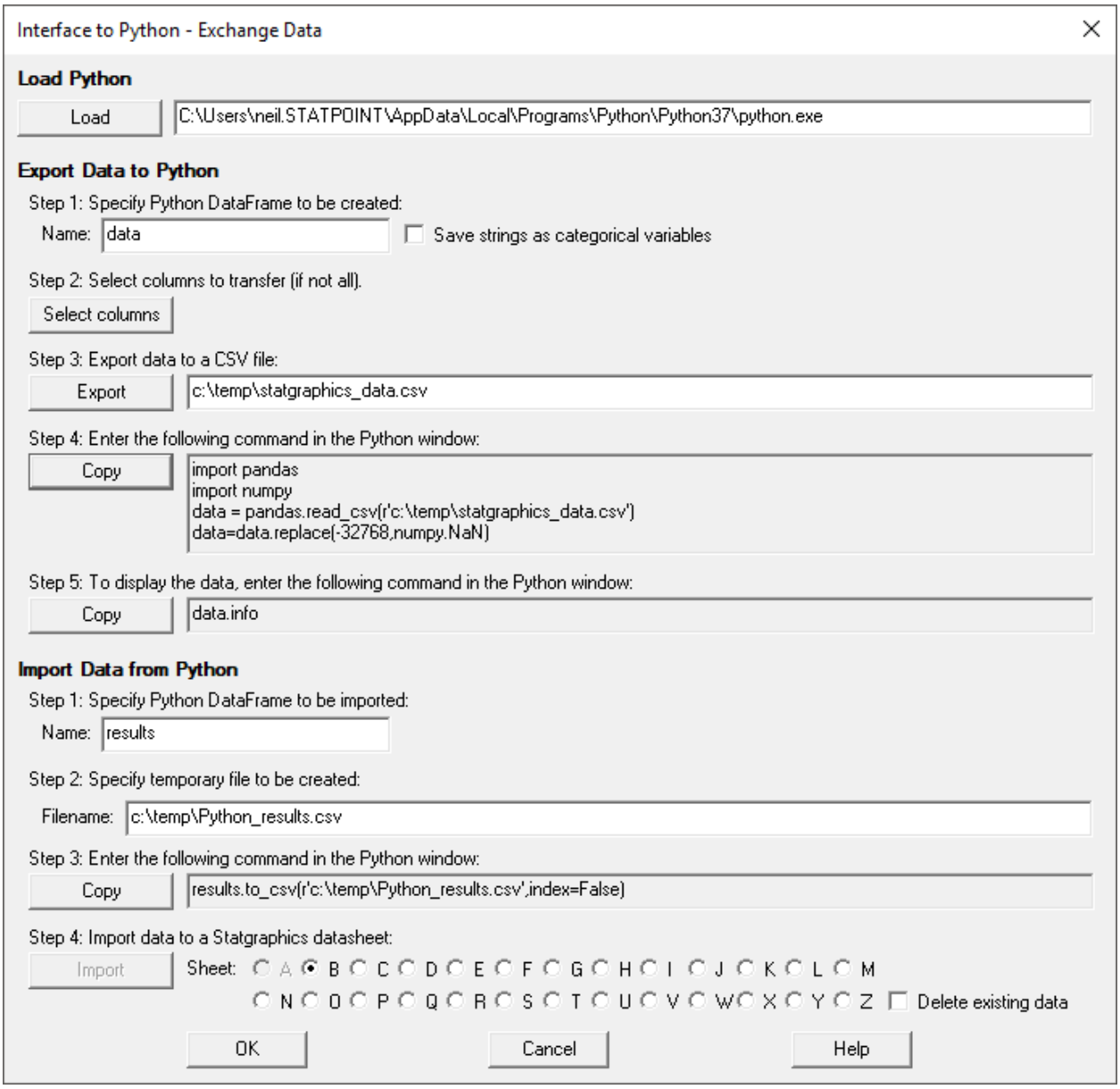
Installation and Configuration

Statgraphics 19 contains a menu item that assists users in installing Python and the required libraries. Python is available for download free of charge on the Internet.



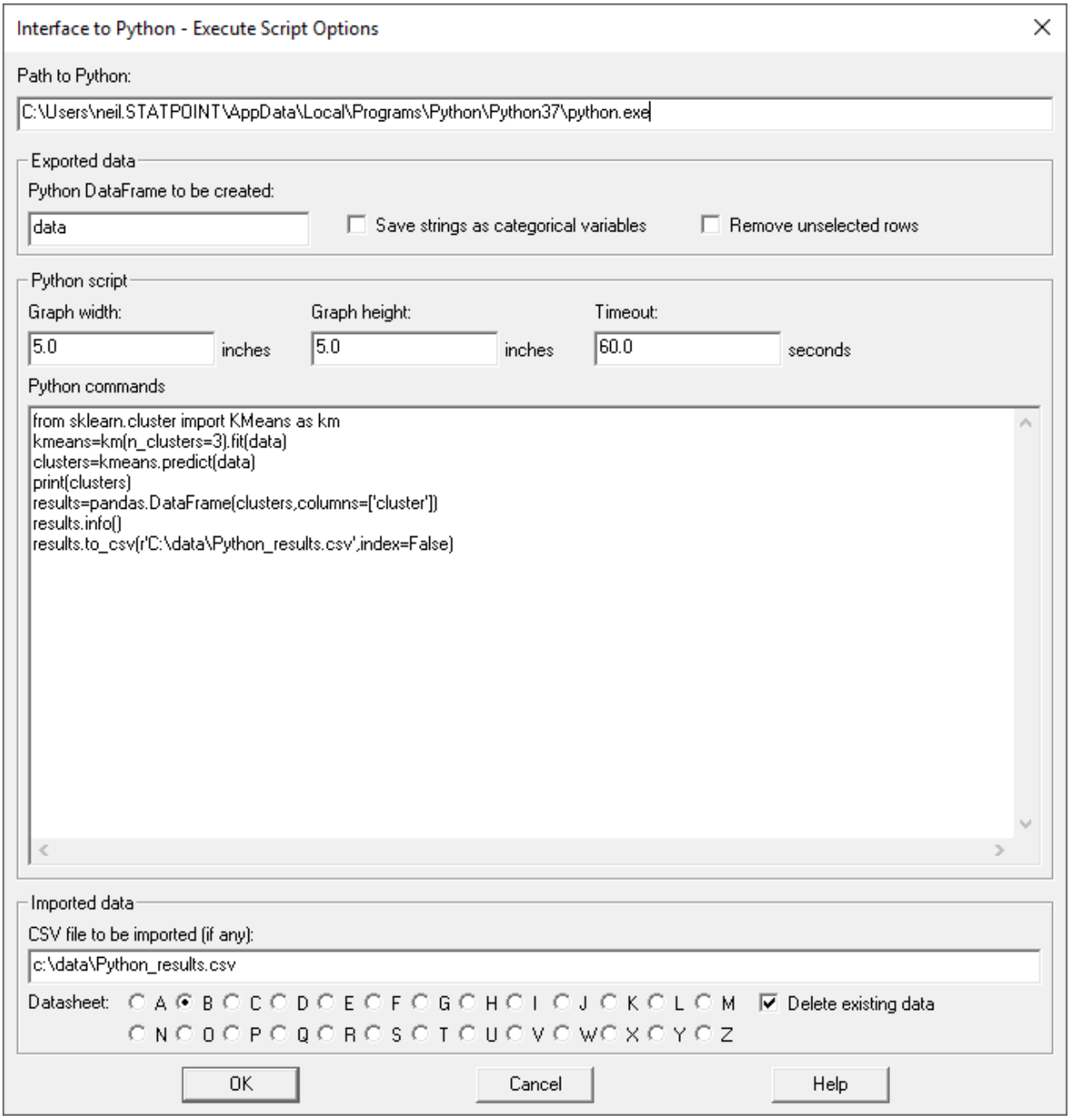
Data Exchange

Data contained in the Statgraphics DataBook may be easily passed to Python. A dialog box is provided to specify the data variables to be passed and the name of the CSV file used to hold the data. Pressing *Export* creates the file and generates the Python commands needed to read the data. The dialog box also assists users in importing results generated by Python by placing them in a Python DataFrame and reading that data into a Statgraphics 19 datasheet.



Executing Python Scripts

As with R, Statgraphics 19 users may create Python scripts, store them in a StatFolio, and execute them at a later time. Data may be passed to Python before the script is executed and results imported after the commands are executed. The script shown below passes data to Python, runs the K-Means clustering procedure, and then reads the resulting cluster numbers back into Statgraphics.

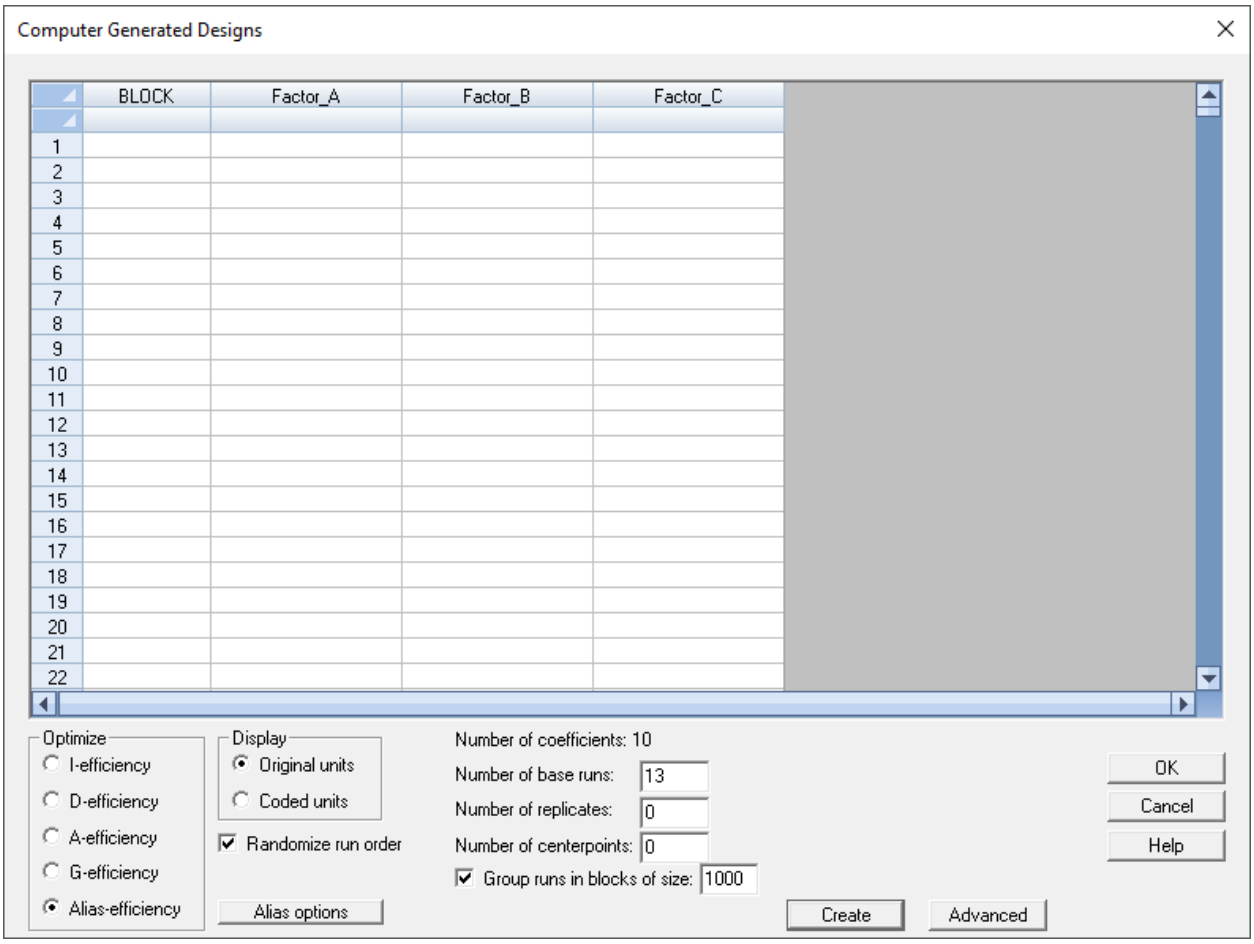


Design of Experiments

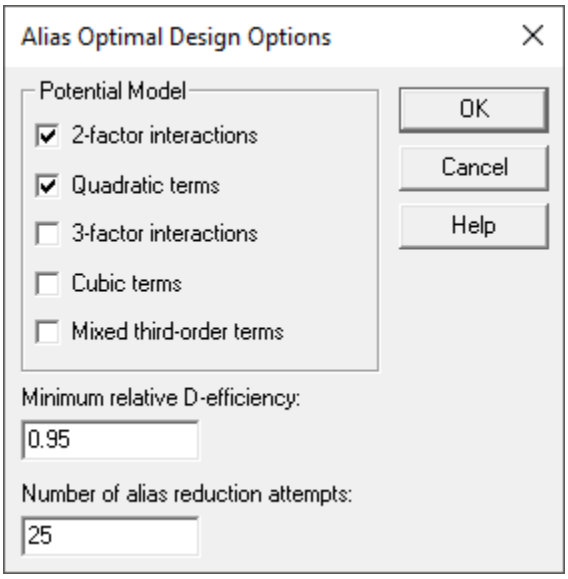
Two significant features have been added to the DOE Wizard: the ability to create *Alias Optimal Designs*, and the ability to enhance previous designs with an optimal set of additional runs.

Alias Optimal Designs

Experimental designs constructed to optimize criteria such as D-efficiency are excellent choices when the existing factors and interactions are known beforehand. In cases when terms not in the assumed model may have a significant effect on the response variables, it is often advantageous to trade a little D-efficiency for a reduction in aliasing in case those effects are active. Statgraphics 19 has added such *Alias Optimal Designs* to the DOE Wizard. To construct such a design, the user first selects a *Computer-Generated Design* and specifies the assumed statistical model. At Step 5, the option to optimize *Alias-efficiency* is then selected:



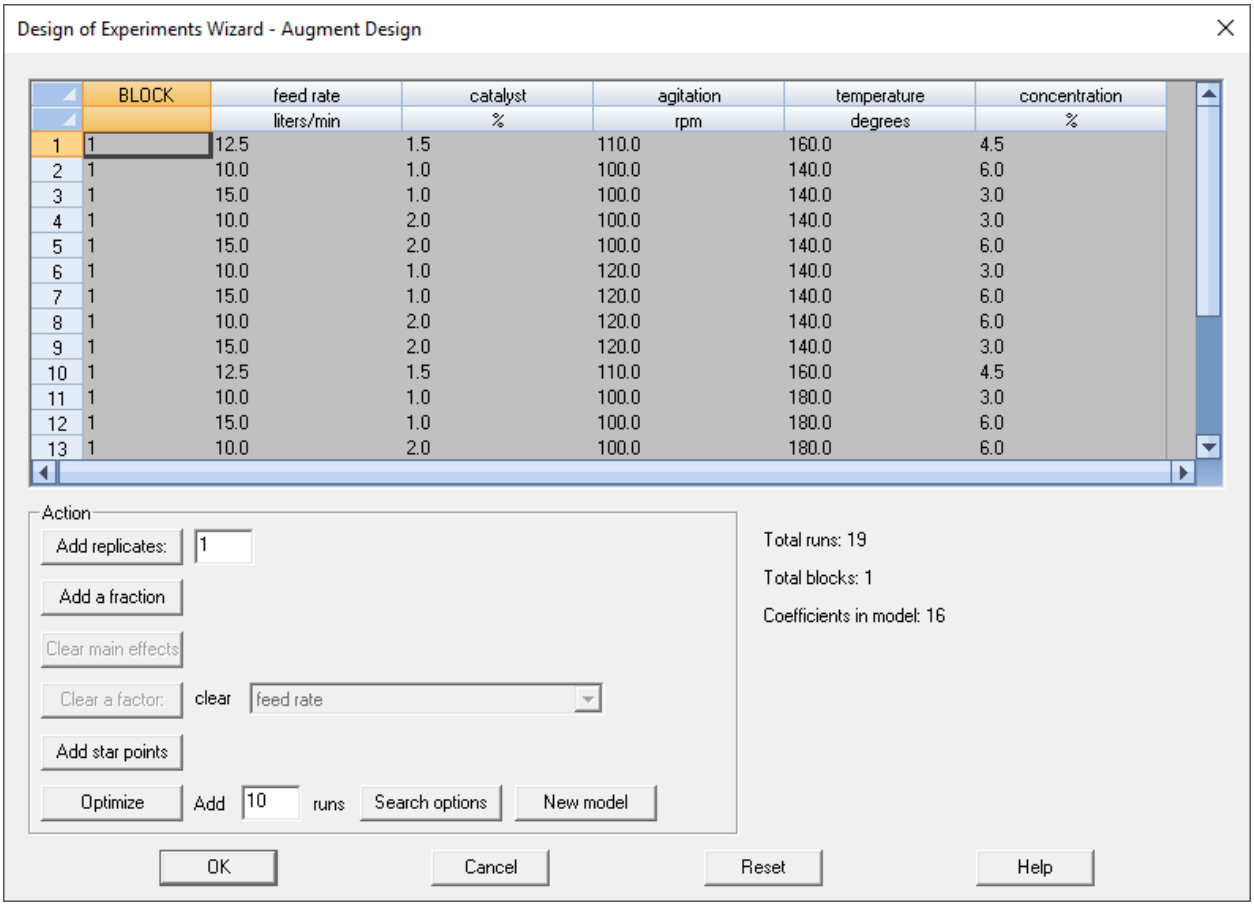
Pressing the *Alias options* button lets the user specify the terms *not* in the current model whose aliasing is to be minimized:



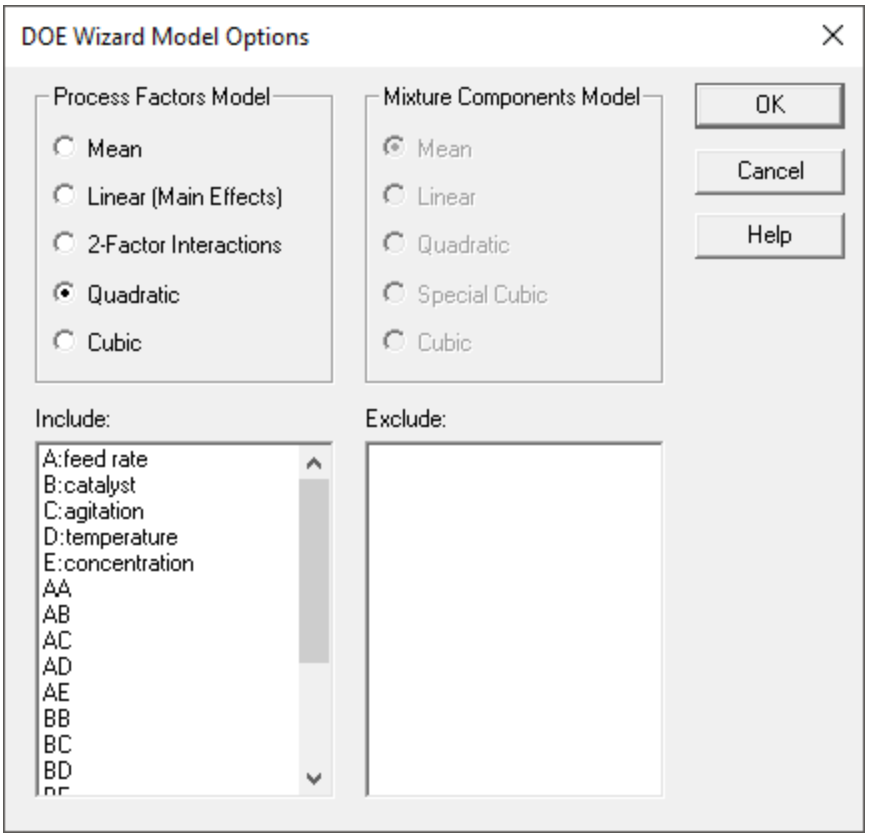
At the same, the maximum allowable tradeoff in D-efficiency to protect against aliasing is specified.

Optimal Augmentation of Existing Design

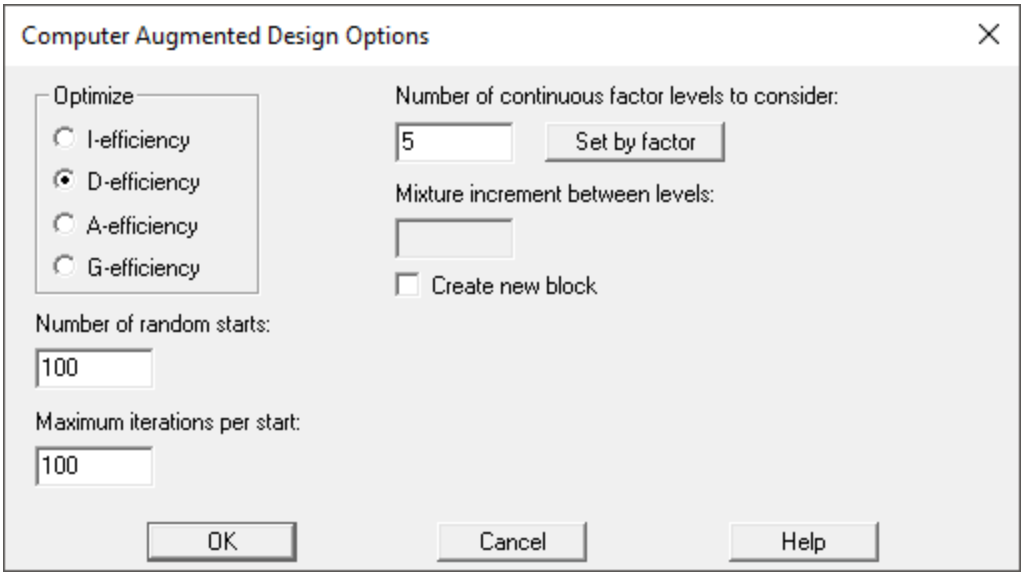
The *Augment Design* step in the DOE Wizard lets the user add runs to an existing design in such a way as to maximize the efficiency of the augmented design.



The user first specifies the number of runs to add and presses the *New model* button to specify the model to be fit with the augmented design:



The *Search options* button controls the criteria used to select an optimal set of additional runs:

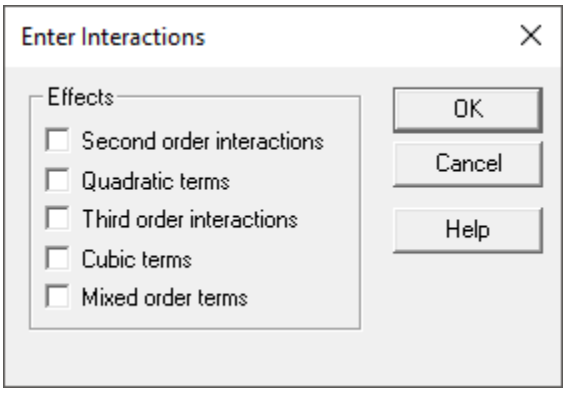


Regression and Analysis of Variance

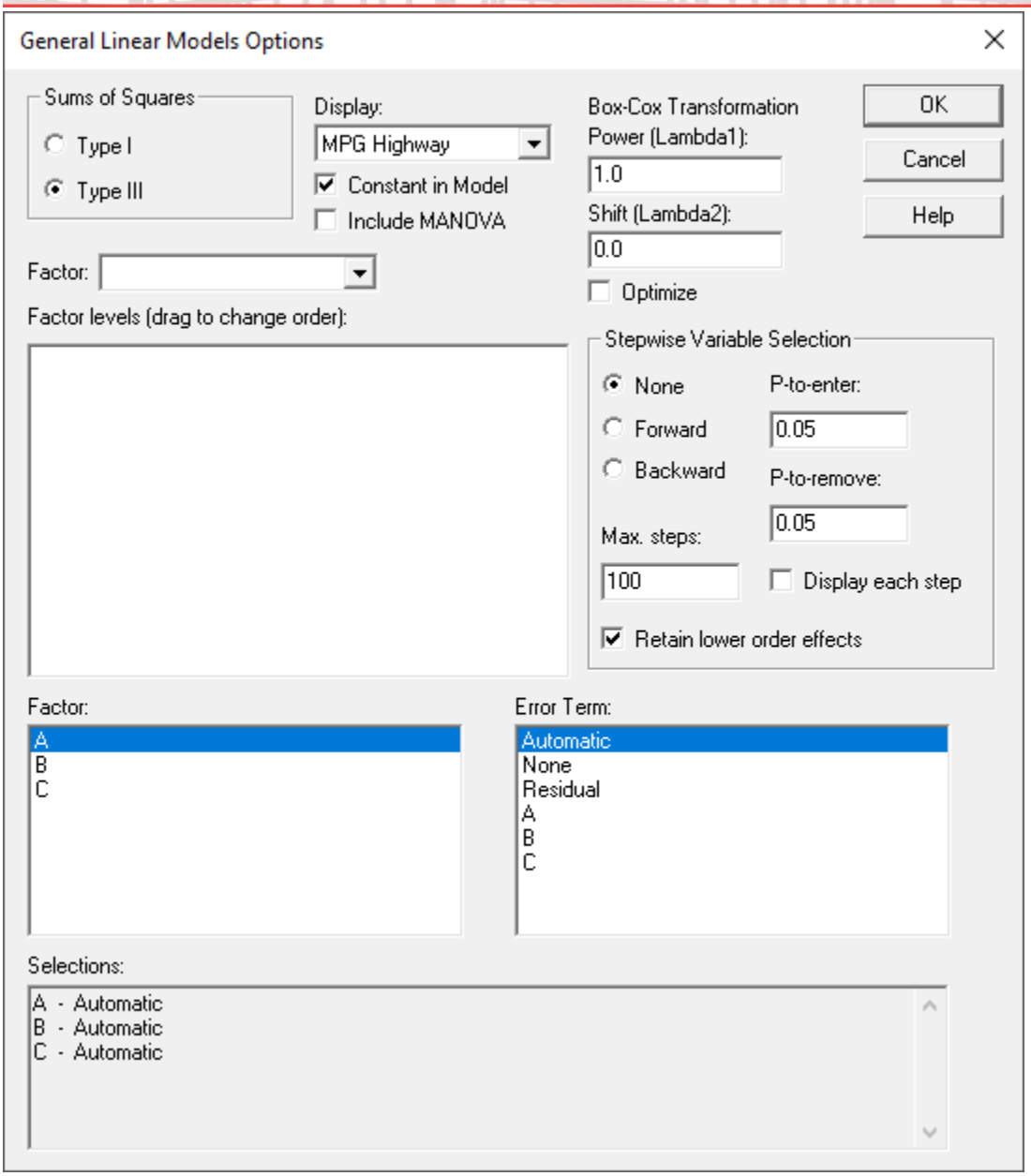
Several new procedures and capabilities have been added for fitting regression models and calculating an analysis of variance.

General Linear Models - Model Specification and Stepwise Selection

The *General Linear Models* procedure has been enhanced in 2 ways: easier specification of models with interactions and polynomial terms, and stepwise selection of variables for models containing both quantitative and categorical factors. When specifying models, a new dialog box has been added which specifies terms to be added to the model specification dialog box:



This greatly reduces the amount of typing required for such models. Second, the *Analysis Options* dialog box lets the user request stepwise selection of factors:



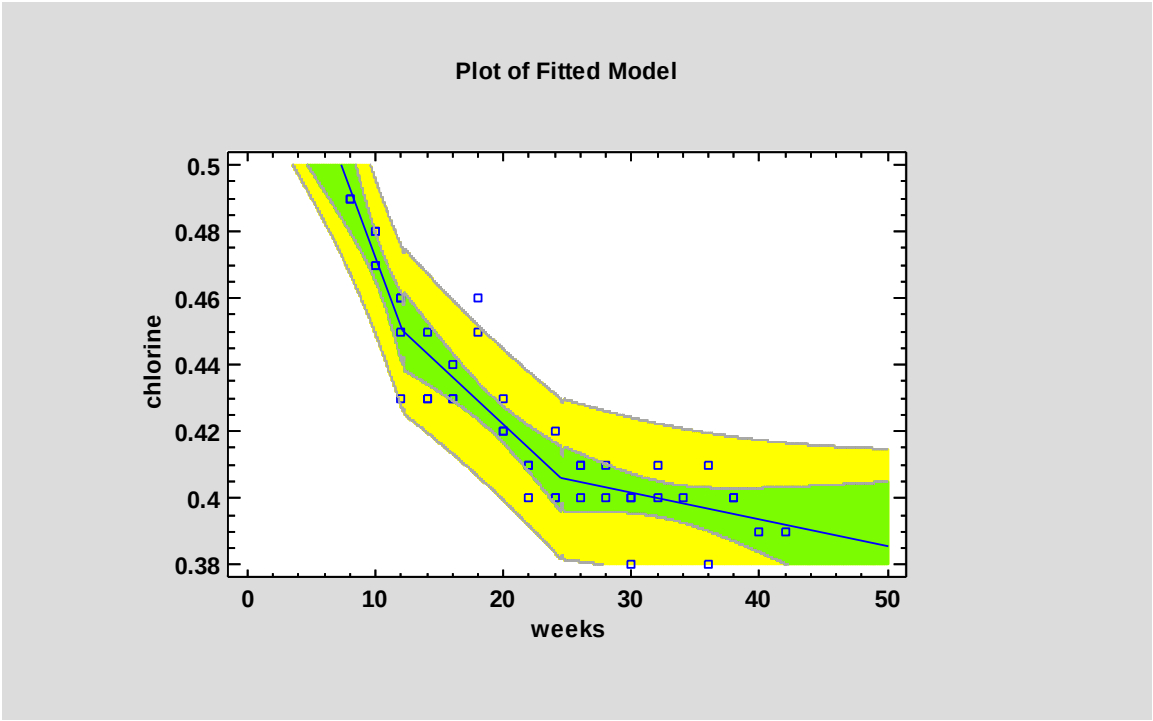
Two features are particularly notable:

1. When a categorical factor is included or excluded, all indicator columns for that factor are treated as a group.

2. The checkbox *Retain lower order effects* prevents consideration of models that contain high order terms but not the corresponding lower order terms. For example, it would not allow a model with the AB interaction unless the main effects of A and B were also included.

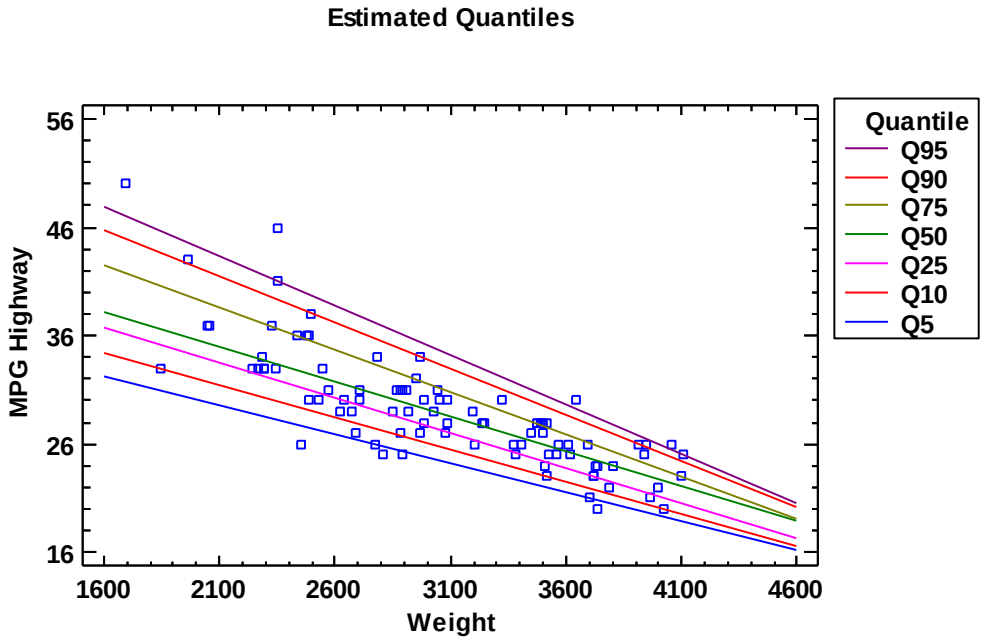
Piecewise Linear Regression

Piecewise linear regression fits a model relating Y and X in which the model consists of 2 or more linear segments with different slopes, and which are continuous where they join. The location of such joints or breakpoints may be either known or unknown.



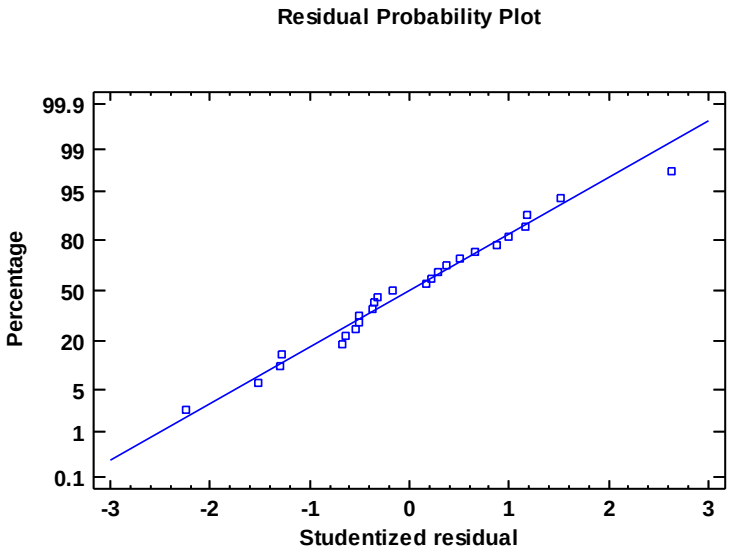
Quantile Regression

Unlike standard multiple regression procedures that construct a model to predict the mean response of Y, quantile regression models are designed to predict a specific quantile. This includes the special case of median regression in which a model is constructed to predict the 50th percentile. Quantile regression is widely used in ecology and econometrics.



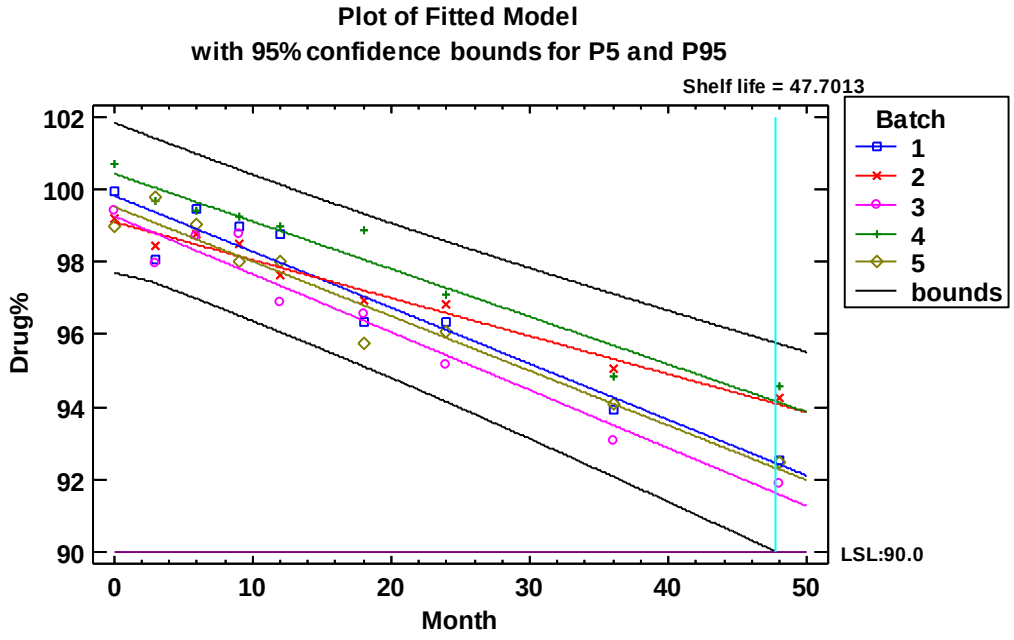
Residual Probability Plots

These plots have been added as an option to several of the modeling procedures.



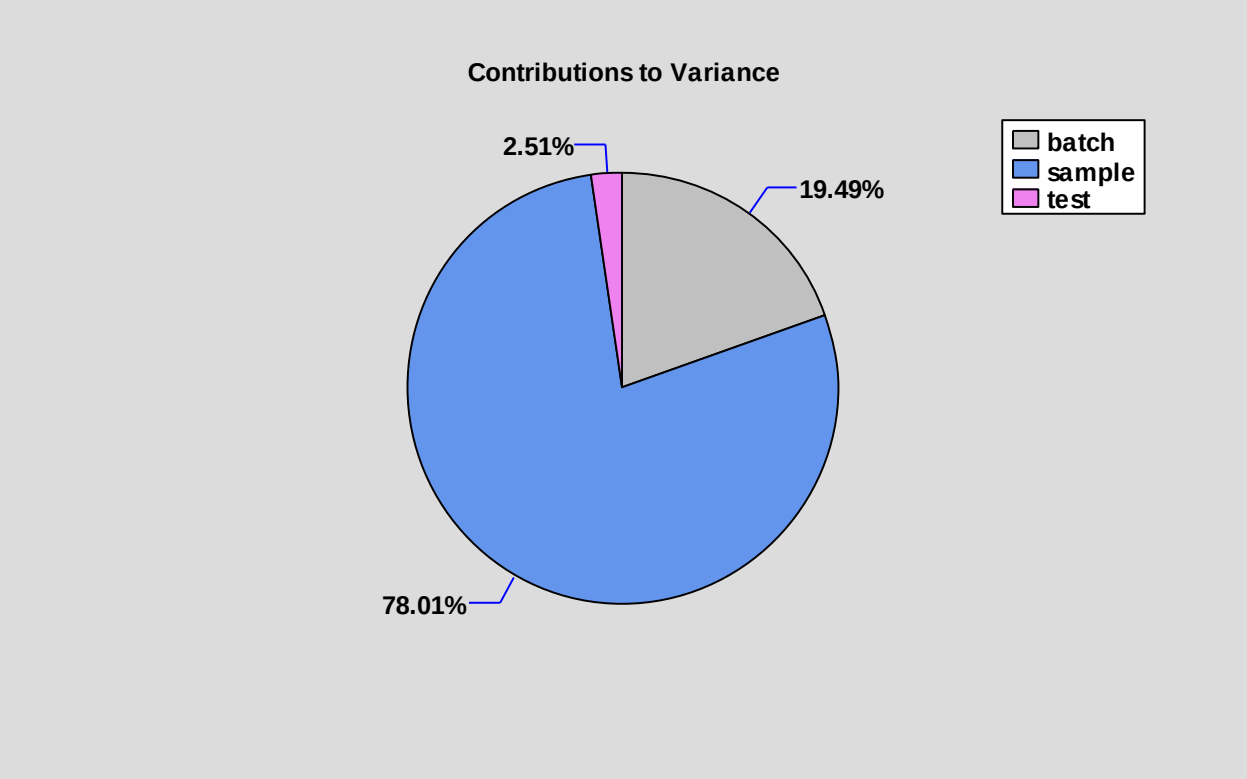
Stability Studies

Stability studies are used to determine acceptable shelf lives for many pharmaceutical, food and other perishable products. A typical study consists of manufacturing several batches of the product and examining their performance at various times after they were produced. The intersection of the 95% confidence bounds for P5 and P95 with specifications for the product is used to determination the product's shelf life.



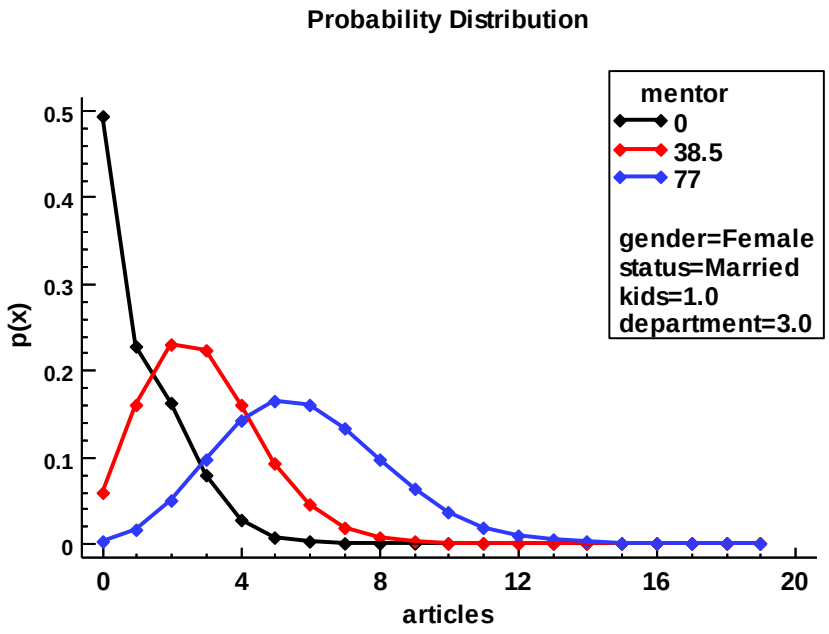
Variance Components Contribution Plot

This plot illustrates the percentage contribution to the overall variance attributable to each of the variance components.



Zero-Inflated Poisson and Negative Binomial Regression

In real life, the distribution of a discrete variable often contains more zeroes than would be predicted by a simple Poisson or negative binomial distribution. In such cases, an additional term that accounts for these extra structural zeroes is added to the distribution. This is important not only when fitting a distribution to a set of identically distributed observations but also when such a variable serves as the dependent variable in a regression model. In the plot below, note the additional zeroes for mentor = 0.



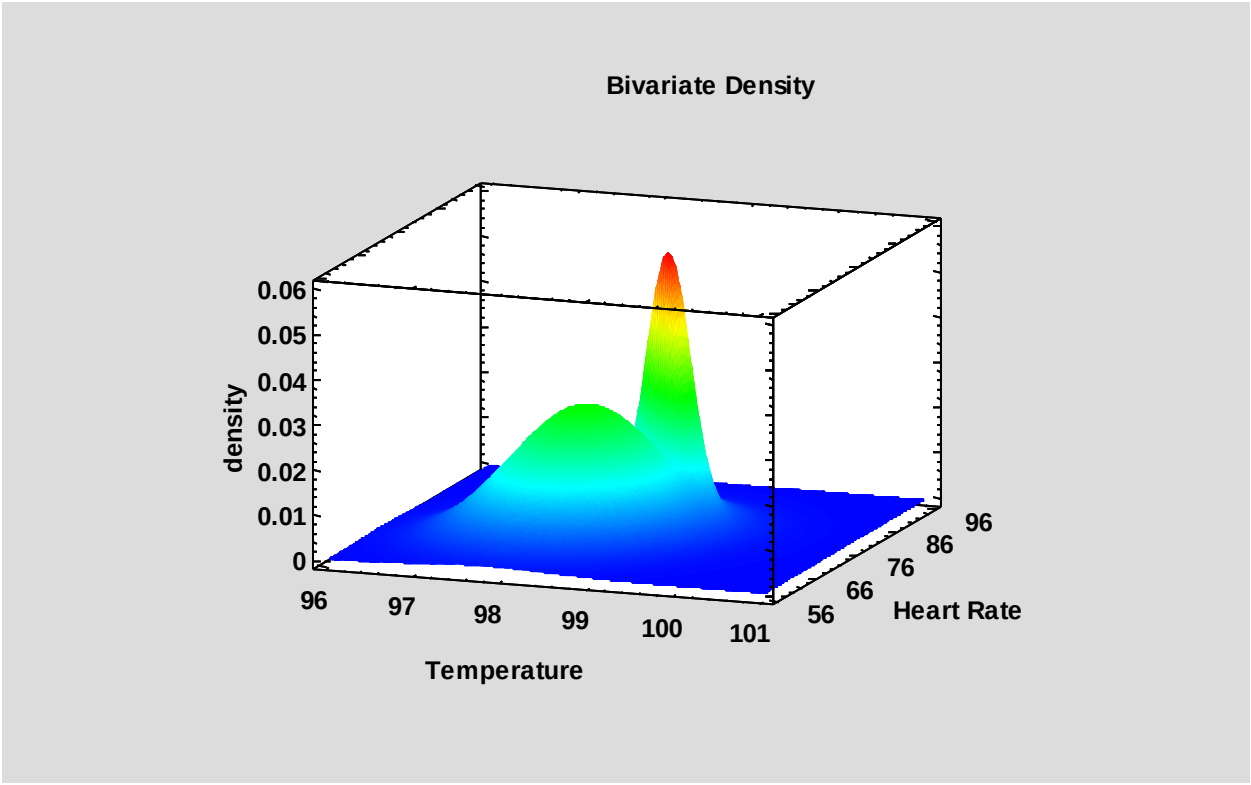
Distribution Fitting

Enhancements to the procedures for fitting distributions include the ability to fit:

1. Bivariate mixture distributions
2. Johnson distributions
3. Univariate mixture distributions
4. Zero-Inflated Poisson and negative binomial distributions

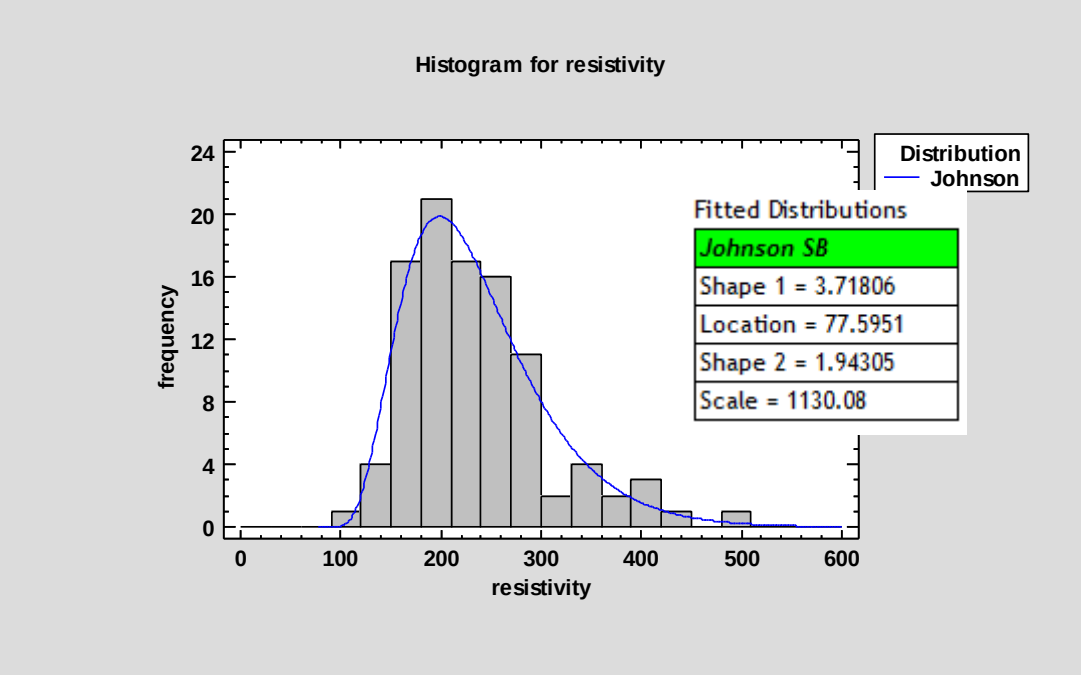
Bivariate Mixture Distributions

Data involving 2 random variables may sometimes be well represented by a mixture of 2 bivariate normal distributions.



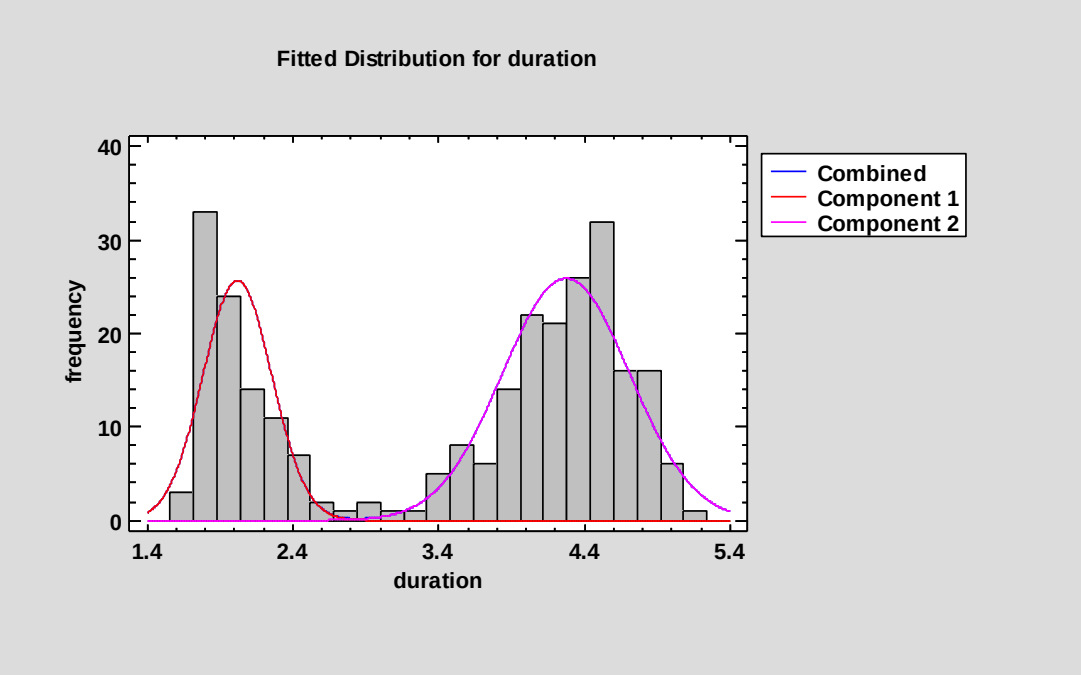
Johnson Distributions

The Johnson family of distributions can match the first four moments of any set of observed measurements. It is helpful in generating an empirical fit to data that do not correspond well to any standard distribution. When fitting data, Statgraphics 19 automatically selects the proper form of Johnson distribution from the SB, SL and SU types.



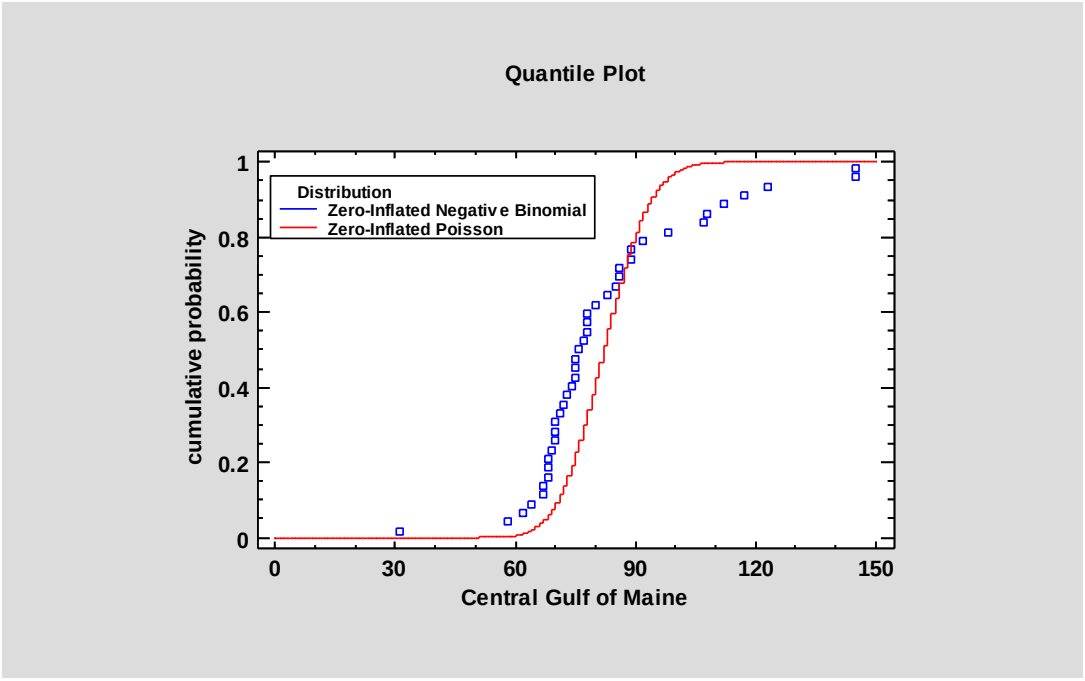
Univariate Mixture Distributions

Real life data are frequently mixtures of more than one distribution. Often, such data can be represented by a mixture of 2 or more normal distributions. Statgraphics 19 can fit such distributions, including estimation of the proportion of observations that come from each component.



Zero-Inflated Poisson and Negative Binomial Distributions

Count data frequently have more zeroes than would normally be expected from distributions such as the Poisson or negative binomial. By adding an additional term to the distribution to represent structural zeroes, a much better fit can often be obtained.



Conformance Analysis

Conformance analysis has been added to the capability analysis procedures for attributes. It is used to estimate the proportion of time that a package containing a fixed number of items will have no more than a specified number of defects. An index called Cpc is also calculated, which compares the estimated probability of conformance with a target value.

Conformance Analysis
Distribution: Binomial
mean percent defective = 1.4

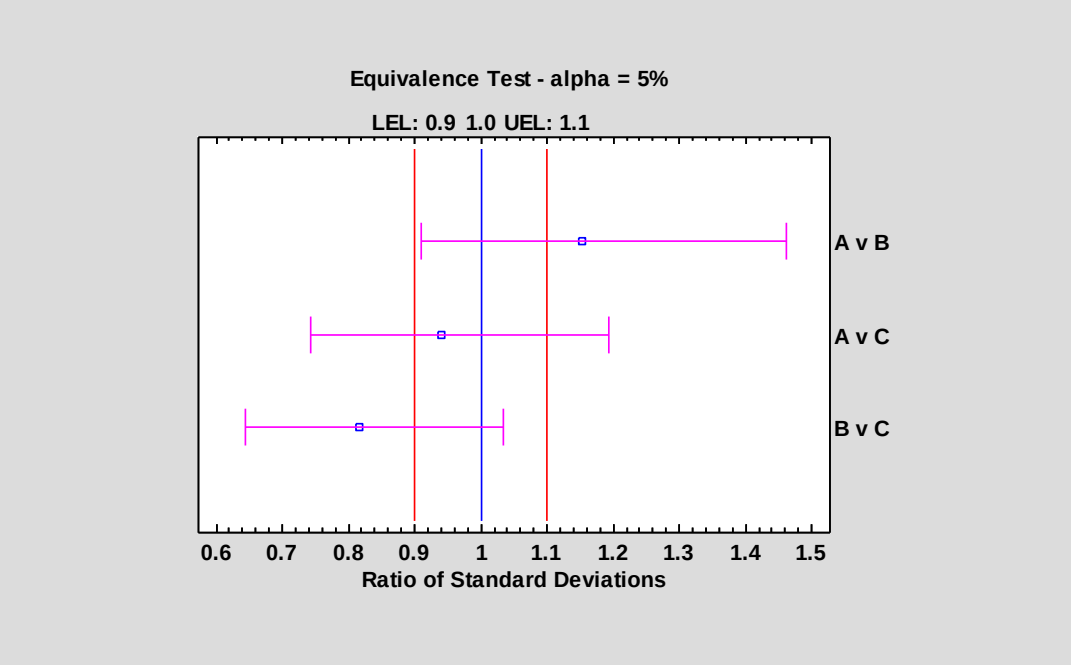
Package size: 500
Maximum allowable nonconforming items: 15
Target conformance: 99.73%

	Estimate	*Lower 95% Limit*	*Upper 95% Limit*
Prob. of conformance	0.997775	0.926574	0.999988
Cpc	1.21359	0.0367719	233.919

The StatAdvisor
The probability of having no more than 15 nonconforming items in a package of 500 is estimated to equal 0.997775. With 95.0% confidence, that probability is somewhere between 0.926574 and 0.999988. The capability index Cpc indicates that the probability of obtaining a nonconforming package is 0.824002 times that allowed (the reciprocal of Cpc).

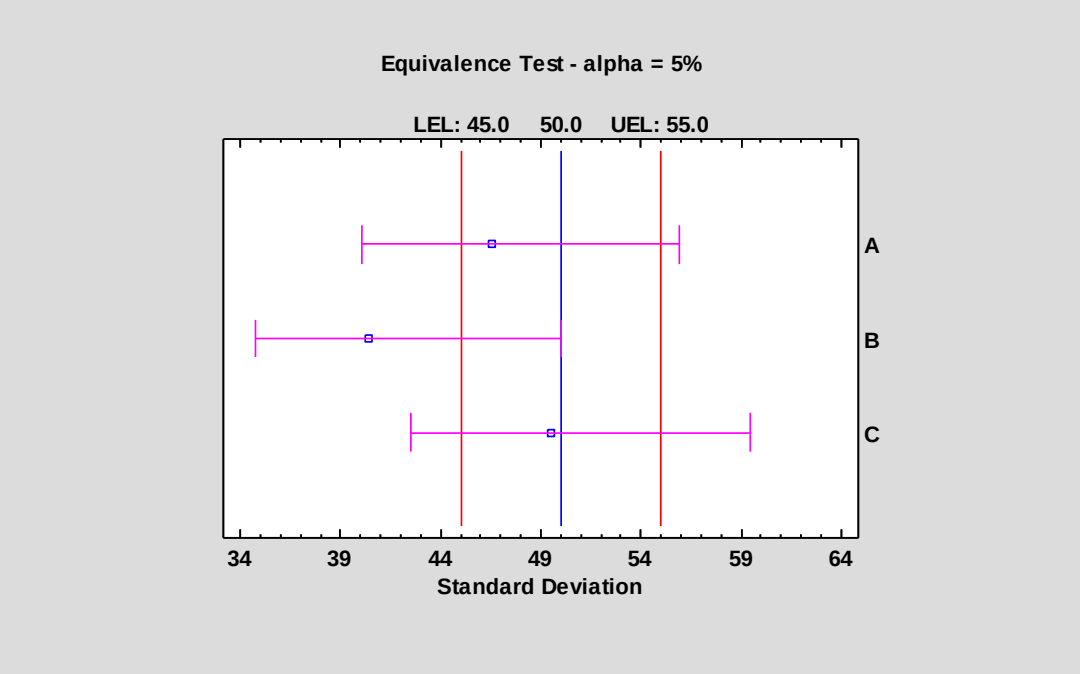
Equivalence and Noninferiority Tests – Comparing 2 Variances

Tests of equivalence and noninferiority were added to demonstrate that the ratio of 2 standard deviations is sufficiently close to 1.



Equivalence and Noninferiority Tests – Comparing a Variance to a Target Value

Tests of equivalence and noninferiority were added to demonstrate that a standard deviation is sufficiently close to a target value.



Mann-Kendall Test

The Mann-Kendall test has been added to the *Runs Chart* procedures to test for the presence of a monotonic trend in a time series.

Mann-Kendall Test
Null hypothesis: no trend
Alternative hypothesis: upward or downward monitonic trend

Test	*Sum*	*Stnd. Error*	*Statistic*	*P-Value*
Mann-Kendall	46	30.7246	1.46463	0.143023

Modified Levene's Test

Levene's test is used to determine whether there are significant differences among the variances of the groups analyzed in a oneway analysis of variance. The *Modified Levene's Test* performs a one-way analysis of variance on the absolute deviations of the data from their respective group medians rather than from their group means as in the unmodified Levene's test.

Variance Check

	Test	*P-Value*
Modified Levene's	0.316515	0.8133

Comparison	*Sigma1*	*Sigma2*	*F-Ratio*	*P-Value*
A / B	9.18753	11.9695	0.589181	0.5018
A / C	9.18753	8.81456	1.08642	0.9157
A / D	9.18753	8.86405	1.07432	0.9271
B / C	11.9695	8.81456	1.84394	0.4381
B / D	11.9695	8.86405	1.82341	0.4464
C / D	8.81456	8.86405	0.988864	0.9886

The StatAdvisor
The statistic displayed in this table tests the null hypothesis that the standard deviations of strength within each of the 4 levels of material is the same. Of particular interest is the P-value. Since the P-value is greater than or equal to 0.05, there is not a statistically significant difference amongst the standard deviations at the 95.0% confidence level.

The table also shows a comparison of the standard deviations for each pair of samples. P-Values below 0.05, of which there are 0, indicate a statistically significant difference between the two sigmas at the 5% significance level.

Unbalanced Gage R&R Studies (with optional additional factors)

A new procedure has been added to perform gage R&R using General Linear Models. It has 2 advantages over earlier procedures:

1. It can handle unbalanced data.
2. Additional factors may be considered beyond just operators, parts and trials.

ANOVA Table

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Operators	927.504	2	463.752	73.0518	0.0000
Parts	22046.4	9	2449.6	385.869	0.0000
Coating	380.992	1	380.992	60.0151	0.0000
Repeatability	679.264	107	6.34826		
Total (Corr.)	24034.2	119			

The StatAdvisor
The ANOVA table divides the total variability of Coating Thickness into several parts. The last part represents the residual error, which corresponds to Repeatability. P-values below 0.05 represent significant sources of variability at the 5.0% significance level.

Wald-Wolfowitz Test

This nonparametric test has been added to the *Two Sample Comparison* procedure. It tests for a significant difference between the distributions generating the 2 samples by combining them and counting the number of runs of values from the same sample.

Wald-Wolfowitz Test
Number of cross-sample ties = 2

	Minimum Statistic	*Minimum P-Value*	*Maximum Statistic*	*Maximum P-Value*	*Average Statistic*	*Average P-Value*
Exact test (number of runs)	6	0.0007	6	0.0007	6.0	0.0007
Large sample z	-3.12406	0.0009	-3.12406	0.0009	-3.12406	0.0009
Large sample z (with cont. corr.)	-2.93662	0.0017	-2.93662	0.0017	-2.93662	0.0017

The StatAdvisor
This option runs a Wald-Wolfowitz test to compare the distributions of the two samples. This test is performed by sorting the combined data from smallest to largest, determining which sample each value came from, and counting the number of runs in the sequence of sample numbers. In this case, the number of runs equals 6. Of particular interest is the P-value for the test. Based on the exact test, P = 0.0007. Since the P-value is less than 0.05, there is a statistically significant difference between the two distributions at the 95.0% confidence level.

Machine Learning

Two new procedures have been added to implement machine learning algorithms.

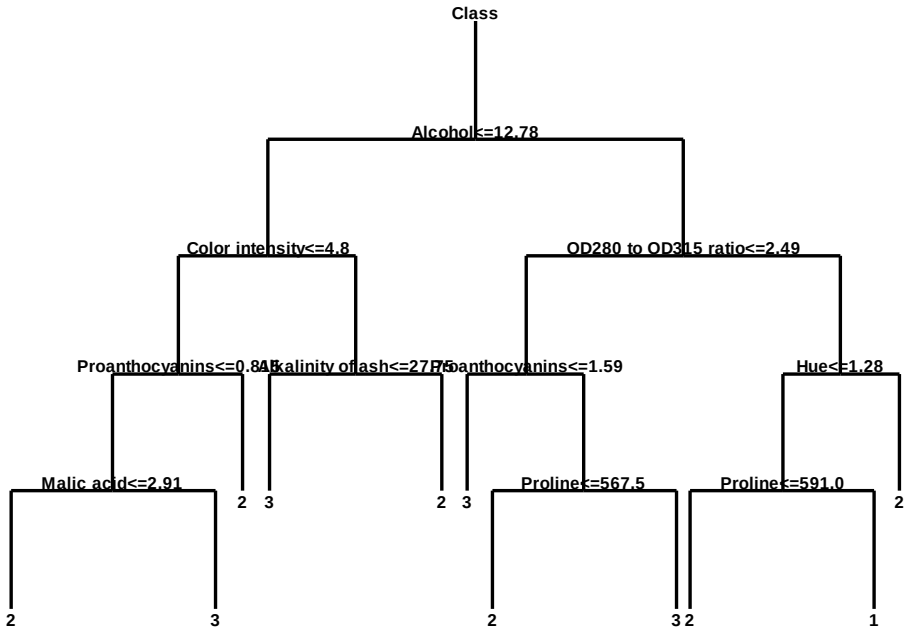
Decision Forests

The *Decision Forests* procedure implements a machine-learning process to predict observations from data. It creates models of 2 forms:

1. *Classification models* that divide observations into groups based on their observed characteristics.

2. *Regression models* that predict the value of a dependent variable.

The analysis constructs a large number of decision trees and averages the predictions made by those trees. Many trees are constructed using a procedure similar to that of *Classification and Regression Trees,* with randomized node optimization and "bagging".

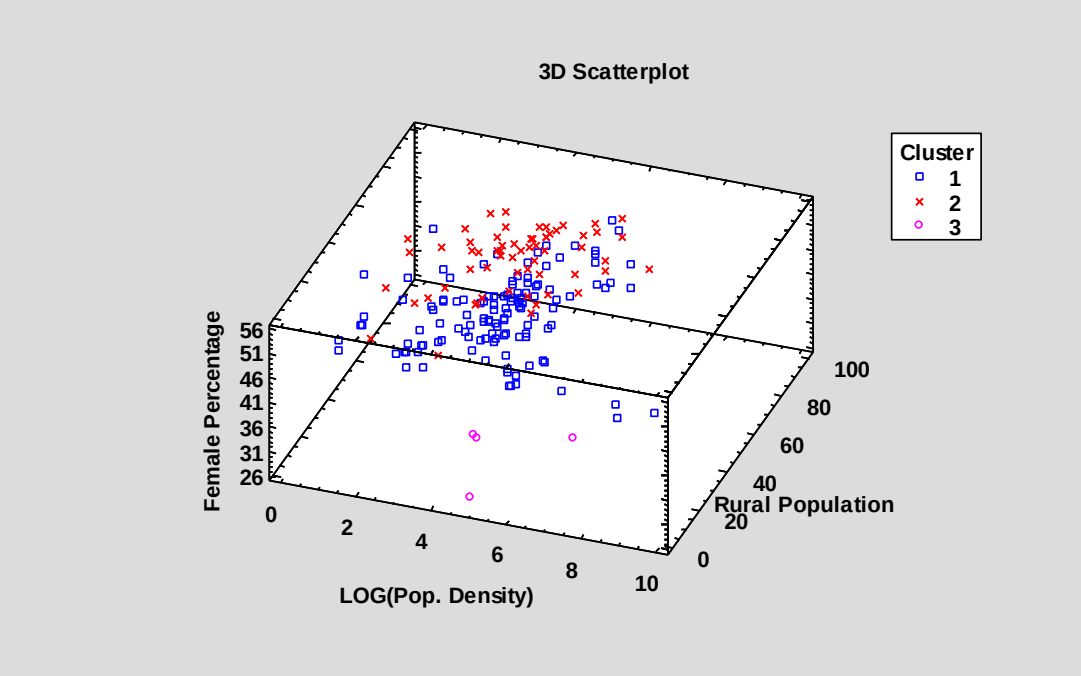


K-Means Clustering

The *K-Means Clustering* procedure implements a machine-learning process to create groups or clusters of multivariate quantitative variables. Clusters are created by grouping observations that are close together in the space of the input variables. Unlike the procedure included in earlier versions, the initial seeds do not have to be specified by the user.

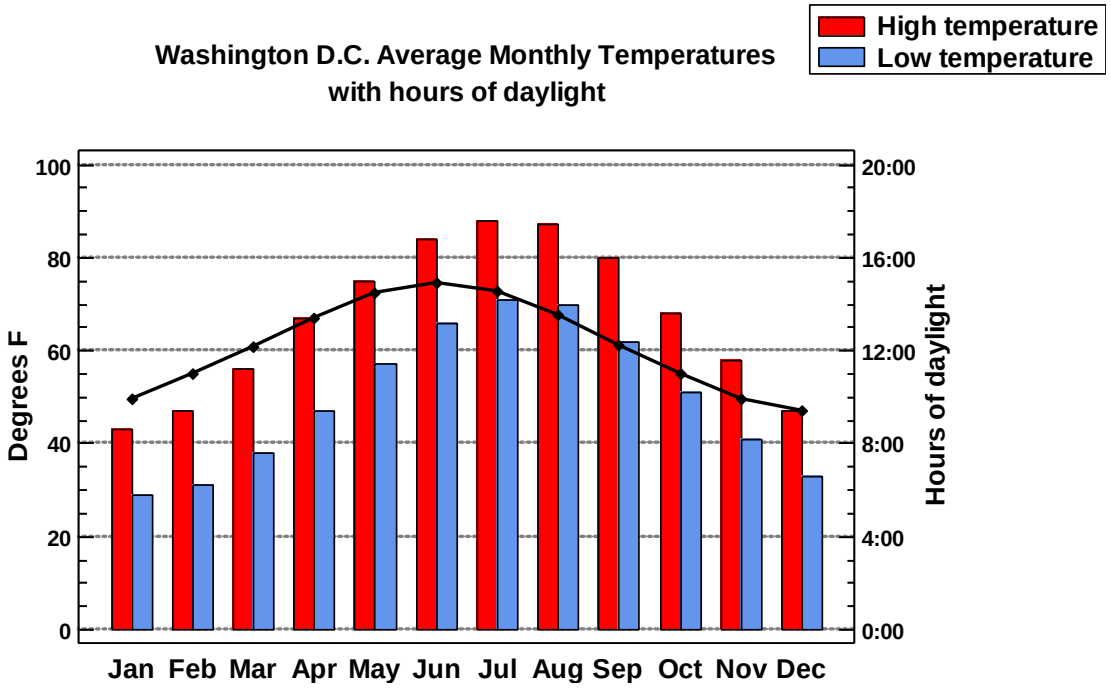
Cluster Summary

Cluster	*Members*	*Percent*
1	123	69.10
2	61	34.27
3	4	2.25



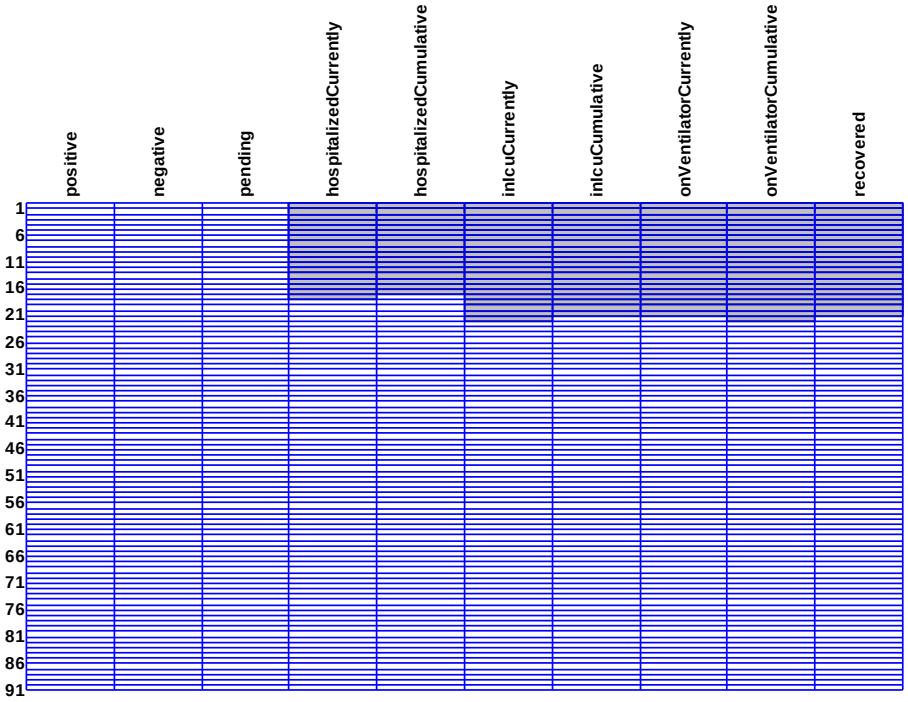
Barchart with Added Line

A line can be superimposed on vertical, clustered barcharts. This feature is available for both simple and multiple barcharts.



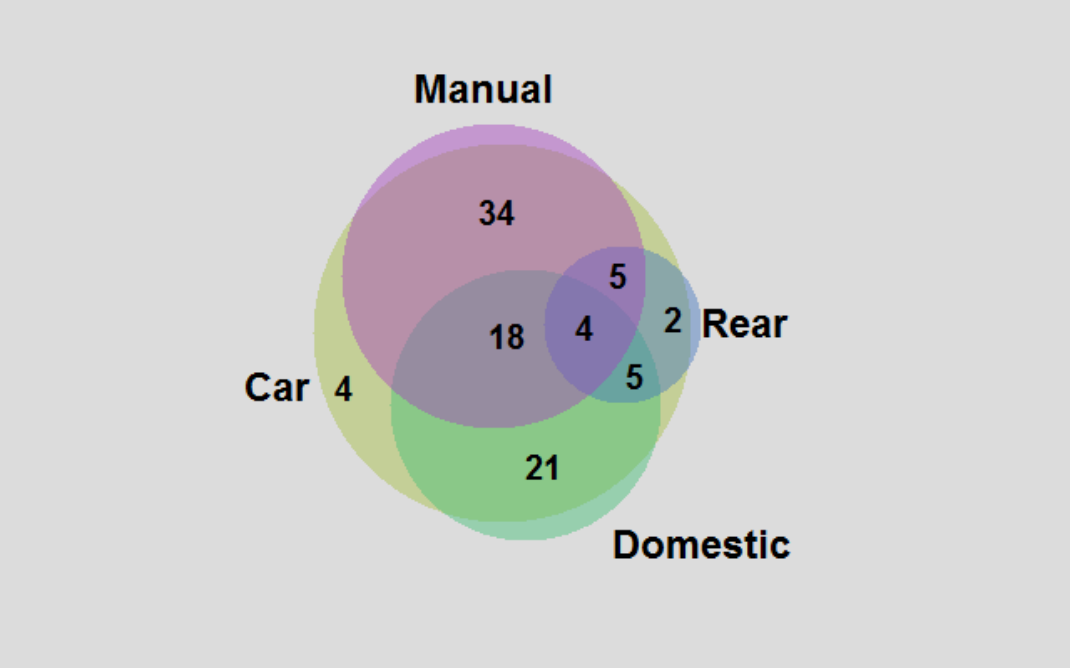
Missing Data Plot

A new plot has been added to the *Data Viewer* to help display the location of missing values in a data file. Shaded cells indicate the location of the missing data.



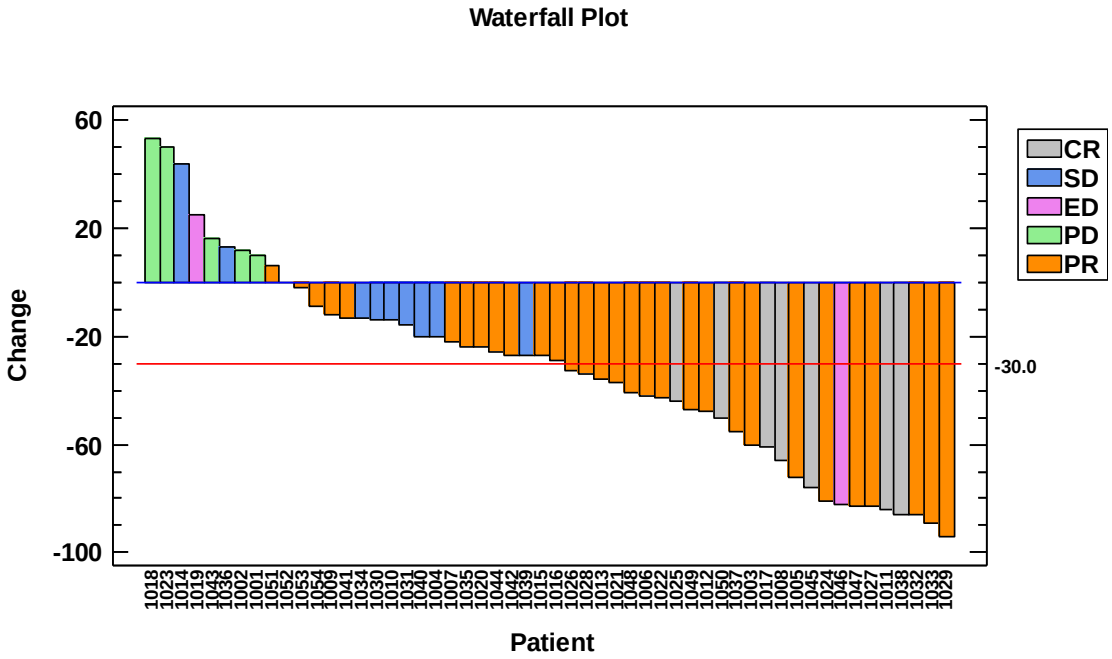
Venn and Euler Diagrams

The Venn and Euler Diagrams procedure creates diagrams that display the relative frequency of occurrence of discrete events. They consist of circular regions that represent the frequency of specific events, where the overlap of the circles indicates the simultaneous occurrence of more than one event.

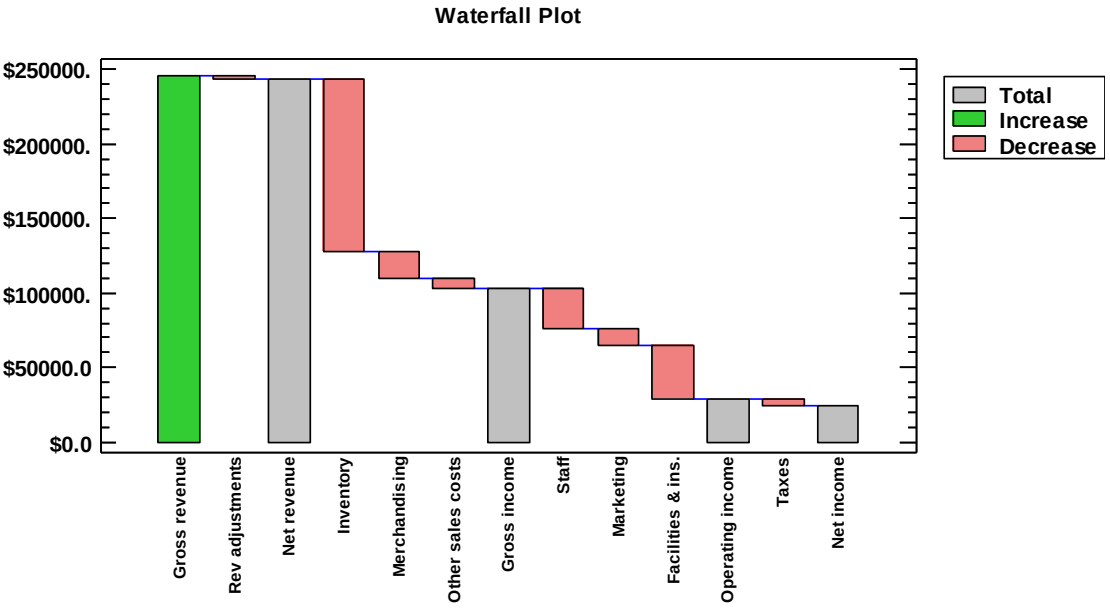


Waterfall Plots

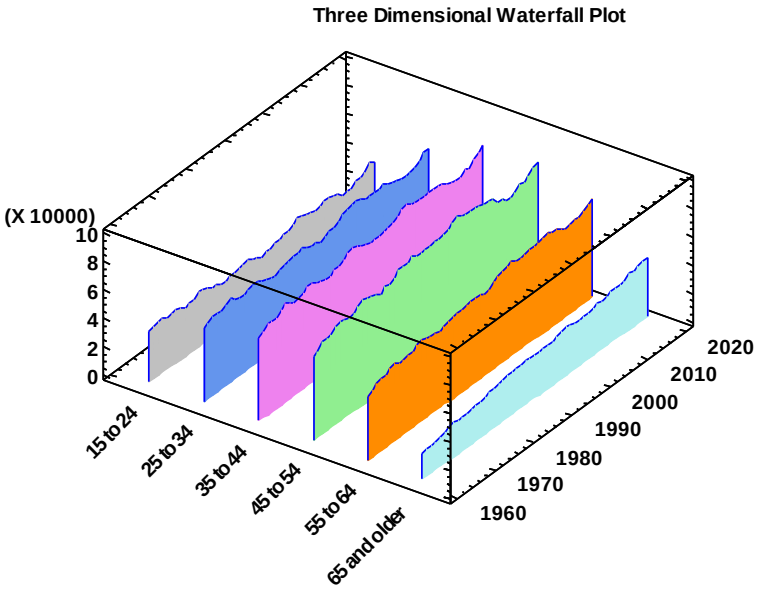
Three new waterfall plots have been added. The first type plots sorted values using a barchart. Colors may be used to indicate a grouping factor.



The second type plots sequential cumulative values, showing increases, decreases and subtotals.



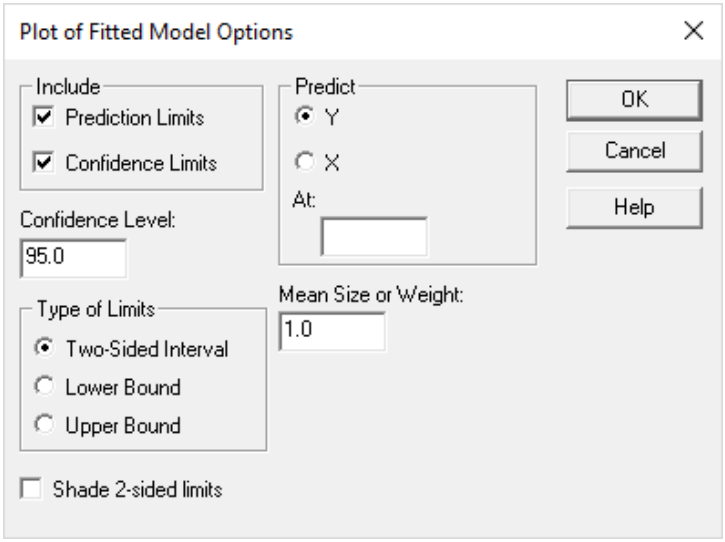
The third type of waterfall plot shows data from a twoway table.



Other Changes

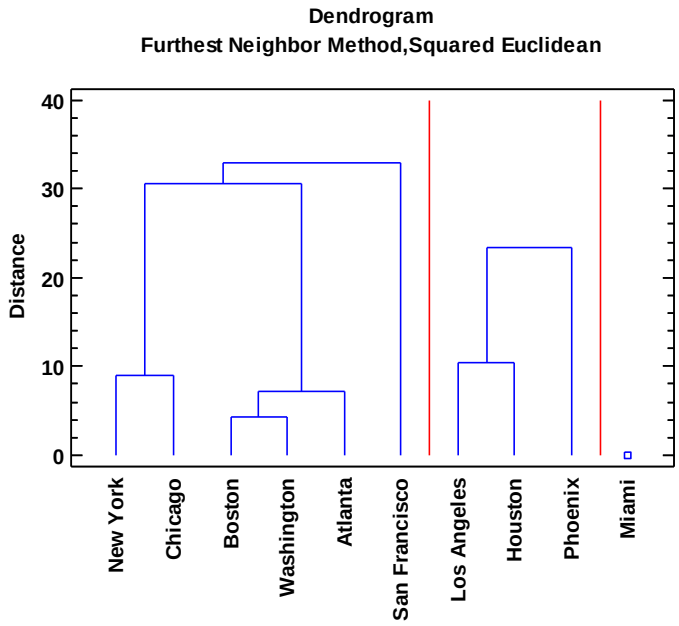
Calibration Models

An option has been added to display one-sided prediction limits:



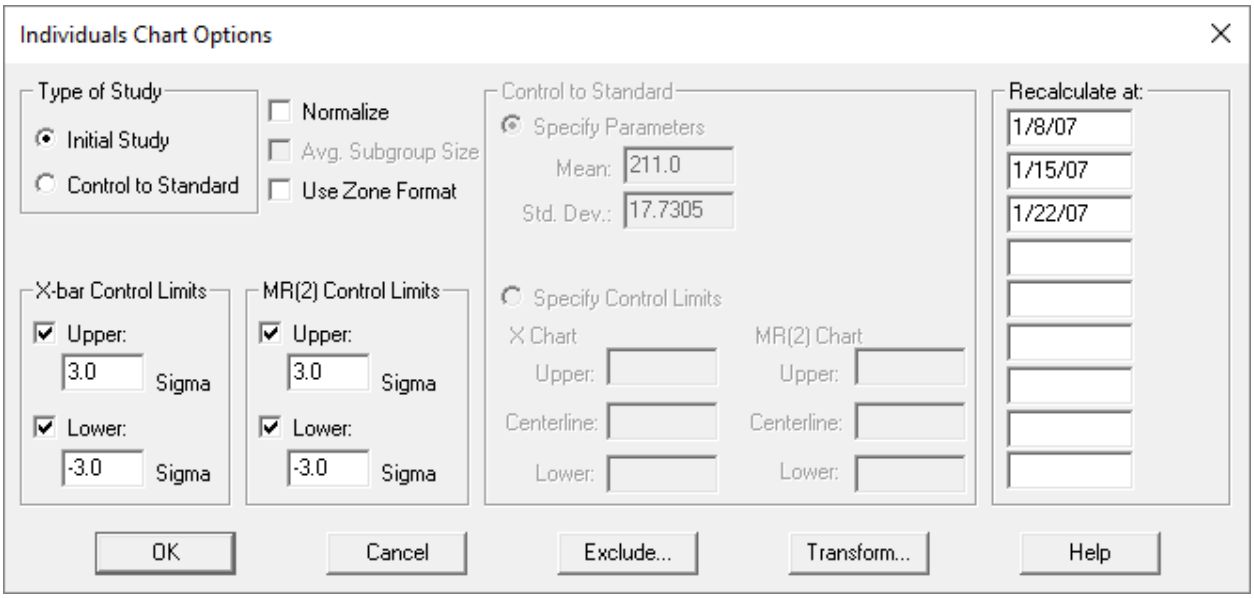
Cluster Analysis

An option has been added to separate clusters on a dendrogram.



Control Charts

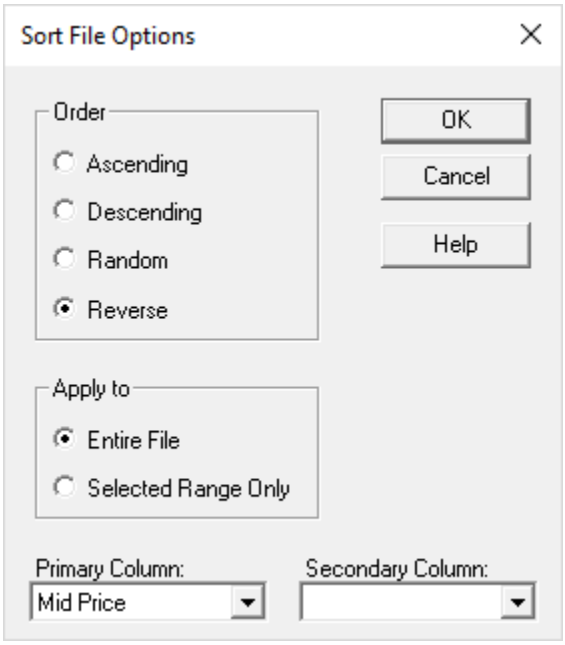
Control limits may now be recalculated at up to 9 locations:



Entries in the *Recalculate at* field may be made using dates or times if the data is in that format.

DataBook – Reverse Sort

A reverse sort option has been added to the *Sort* item on the *Edit* menu:



If selected, it reverses the order of the rows in the current datasheet.

DataBook - Undo

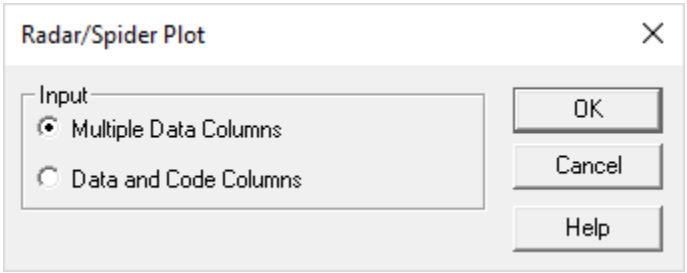
Undo now reverses up to 30 operations performed in the Databook.

Data Import

Direct data import from Minitab project files, SAS transport files, and SPSS portable files has been added. This is in addition to import from standard Minitab, SAS and SPSS data files.

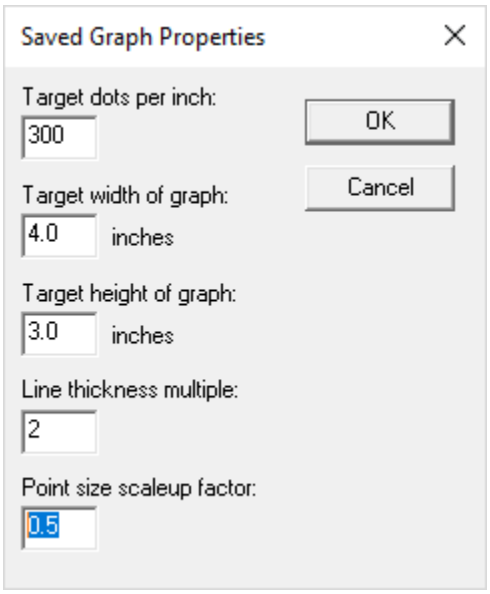
Radar/Spider Plots

An option for inputting data and code columns in addition to multiple data columns has been added:



Save Graph

An option has been added when graphs are saved in image files to increase or decrease the size of all point symbols:



StatGallery

Graphs contained in the StatGallery may now be saved to image files in the same manner as graphs in an analysis window.

Transparent Background on Graphs

The opacity setting for graphs is now retained when graphs are copied to the Windows clipboard for insertion into other applications.