



# STATGRAPHICS Centurion XVI ENHANCEMENTS

## DATA MANAGEMENT

- 1. DataBook** - The number of datasheets in the DataBook has been expanded from 10 to 26. This means that a single StatFolio can access up to 26 data sources at one time. It also provides targets for up to 16 response variables in a designed experiment and expanded capabilities for file-splitting operations.
- 2. Variable Names** - Variable names may now contain any character, including parentheses, dashes and other algebraic symbols that were not permitted in earlier versions.
- 3. Data Files** - The default format for STATGRAPHICS data files is now XML, which allows the data to be edited using any text editor. Version XVI is backwards compatible, however, and can still read and write files in the older STATGRAPHICS formats.
- 4. Data Import** - Version XVI can now read data files created by the latest version of Excel (extension .XLSX). It can also import files created by most statistical packages, including SAS, SPSS, Minitab, JMP, Statistica, Systat, Mathcad, Gauss, R, S-Plus, and Stata.
- 5. Recent File List** - The recent file list now shows the eight most recently accessed data files, as well as the eight most recent StatFolios and the eight most recent XML scripts.
- 6. Transposed Rows and Columns** - A new *Copy Transposed* selection has been added to the *Edit* menu that transposes rows and columns when a selection of cells is copied to the clipboard. This makes it easy to transpose rows and columns when needed in order to match the required data structure for a statistical procedure.
- 7. Find and Replace** - An option has been added to find and replace specified strings within a datasheet.
- 8. Data Viewer** - A new procedure has been added to view data files. The procedure produces a quick summary of all variables in the file.

## GRAPHICS

- 1. Pan and Zoom** - A pan and zoom capability has been added that allows users to zoom in on selected portions of any graph and then dynamically pan back and forth in different directions. This is very helpful when the number of observations plotted is large.
- 2. Fonts** - Font sizes have been changed to allow finer control over the size of text on graphs. This allows for smaller fonts to be used to label points.
- 3. Dynamic Rotation of 3-D Graphs** - All three-dimensional graphs may now be rotated dynamically using scrollbars bordering the graphics window.
- 4. Dynamic Jittering, Brushing and Smoothing** - The dialog boxes for the jitter, brush and smooth operations have been made dynamic so that users can observe the effects as they interact with the dialog box controls.
- 5. Logarithmic Scaling** - A new format for logarithmic scaling now supports log scales with tickmarks that are not all powers of ten.
- 6. Axis Tickmarks** - An option has been added to suppress the gap between the axes and the first major tickmark. Minor tickmarks can also be suppressed.
- 7. Dynamic Response Surface Exploration** - 3-D response surface plots can now be explored dynamically by interactively changing the levels of one or more variables. This greatly facilitates the interpretation of effects in designed experiments, as the presence or absence of interactions is easily visualized. In addition, if the program is asked to follow the path of steepest ascent, the plots will change dynamically as the values of the factors are slowly changed.
- 8. Graphics Profile Designer** - A new option has been added to the *Tools* menu to make it easier to set and save all graphics options. Sample 2-D and 3-D plots are created showing all options, allowing users to create and save profiles that cover all types of graphs in the program.



- 9. Point Labels** - The *Point Identification* button has expanded options for automatically labeling points on any scatterplot. For plots such as control charts, labeling can be restricted to unusual points only. Labels can also be added only when a point is clicked on with the mouse.
- 10. 3-D Contour Plots** - Three-dimensional contour plots display contours for three factors simultaneously. They can be manipulated dynamically using the response surface explorer.
- 11. 3-D Mesh Plot** - Three-dimensional mesh plots display the value of a response throughout a three dimensional region.
- 12. Named Web Colors** - Colors may now be selected using standard Web names.

## SYSTEM OPERATION

- 1. XML Scripts** - The STATGRAPHICS *XML Scripting Language* allows users to execute STATGRAPHICS procedures without using menus or dialog boxes. Using the scripting language, system globals can be set, data files can be accessed and statistical procedures can be invoked. Scripts developed with the XML scripting language may be executed by:
  - STATGRAPHICS Centurion Version XVI or later; or
  - The STATGRAPHICS.Net Web Services.
- 2. System Settings** - It is now possible to export, import or restore default system settings.
- 3. Procedure Operations** - When procedures are selected from the menu, three dialog boxes are displayed before the analysis window is created:
  - Data Input dialog box;
  - Analysis Options dialog box; and
  - Tables and Graphs dialog box.
- 4. Procedure Tables and Graphs** - Users may save their desired default tables and graphs for each procedure.
- 5. Repeat Analysis "BY"** - A new option has been added to the *Edit* menu that allows users to run an analysis for each unique value contained in a "BY" variable.

## STATISTICAL PROCESS CONTROL

- 1. StatFolio Alerts** - Alerts can now be generated automatically whenever one of several events occurs:
  - A plotted point falls beyond the upper or lower control limit on a control chart;
  - An unusual run is observed on a control chart; or
  - A calculated capability index falls above or below a specified value.Alerts can be in the form of dialog box messages, audio signals or e-mails sent to specified addresses. A log of all alerts is kept for later review.
- 2. Control Chart Scaling** - Additional fields have been added to all control chart data input dialog boxes allowing users to specify a variable containing locations along the x axis. This provides for improved scaling of dates and times.
- 3. Specification Limits** - Control charts now allow users to add specification limits to the charts.
- 4. Six Sigma Calculator** - The functionality of the Six Sigma Calculator has been expanded. In addition to converting among various quality metrics, it also creates a plot illustrating the estimated area within the specification limits.
- 5. Normal Tolerance Limits** - When normalizing transformations are necessary, the transformations will be automatically inverted after calculating normal tolerance limits.

## DESIGN OF EXPERIMENTS

- 1. DOE Wizard** - The *Experimental Design* section of STATGRAPHICS contains a new wizard that assists users in constructing and analyzing designed experiments. It guides the user through 12 important steps.  
The first seven steps are executed before the experiment is run:  
Step 1: Defining the response variables;  
Step 2: Defining the experimental factors;  
Step 3: Selecting the appropriate experimental design;



Step 4: Defining the model to be fit to the data;  
Step 5: Selecting an optimal subset of the experimental runs (if creating a D-optimal design);  
Step 6: Evaluating the design; and  
Step 7: Saving the experiment that has been created.

The final five steps are executed after the experiment has been performed:

Step 8: Analyzing the results by constructing a statistical model for each response variable;  
Step 9: Finding the setting of the experimental factors that optimize the responses;  
Step 10: Saving the results;  
Step 11: Augmenting the design if necessary by adding additional runs; and  
Step 12: Extrapolating the models beyond the experimental region to search for locations that may yield even better results.

**2. Combined designs for quantitative and categorical factors** - The DOE Wizard permits the construction and analysis of designs that include both quantitative and categorical factors.

**3. Combined designs for process and mixture variables** - The DOE Wizard permits the construction and analysis of designs that include both process factors and mixture components.

**4. Robust parameter designs** - The DOE Wizard can create experimental designs for use in robust parameter design (RPD). Such experiments include both controllable factors and noise factors. The goal of RPD is to find levels of the controllable factors where the response variables are relatively insensitive to changes in the noise factors. Using the STATGRAPHICS Centurion XVI DOE Wizard, robust parameter designs can be constructed in two ways:

- *Crossed approach* – Two separate designs are created, one for the controllable factors and one for the noise factors.

These two designs are then merged by creating runs with all combinations of runs from both designs. This is the method first suggested by Genichi Taguchi and is described in the STATGRAPHICS document titled *DOE Wizard – Inner/Outer Arrays*.

- *Combined approach* – Both the controllable factors and the noise factors can be studied in a single design. This approach is described by Myers, Montgomery and Anderson-Cook (2009) and has several advantages, including fewer total runs and more insight into the effects of the factors on both the mean and variance of the response.

**5. Integrated multiple response optimization** - In order to find a combination of the experimental factors that provides a good result for multiple response variables, the DOE Wizard uses the concept of *desirability functions*. The DOE Wizard will search the experimental region for combinations that yield the highest desirability.

**6. Extrapolation along path of steepest ascent** - The statistical models can be extrapolated outside of the experimental region in order to suggest the best direction in which to look for potential improvements in the desirability of the response variables.

**7. Prediction variance plots** - These plots show how the standard error of the predicted response varies throughout the experimental region. The more constant the variance is, the more preferable the design.

**8. Variance dispersion graphs** - These graphs illustrate the change in the minimum, average and maximum scaled prediction variance as a function of the distance from the center of the experimental region.

**9. Fraction of design space plots** - FDS plots show the fraction of the design space where the scaled prediction variance is less than different values. A mostly level plot indicates that the variance is low throughout much of the experimental region.

## NEW STATISTICAL PROCEDURES

**1. Correspondence Analysis** - The Correspondence Analysis procedure creates a map of the rows and columns in a two-way contingency table for the purpose of providing insights into the relationships among the categories of the row and/or column variables. Often, no more than two or three dimensions are needed to display most of the variability or “inertia” in the table. An important part of the output is a correspondence map on which the distance between two categories is a measure of their similarity.



**2. Multiple Correspondence Analysis** - The Multiple Correspondence Analysis procedure creates a map of the associations among categories of two or more variables. It generates a map similar to that of the *Correspondence Analysis* procedure.

However, unlike that procedure which compares categories of each variable separately, this procedure is concerned with interrelationships among the variables.

**3. One Dimensional Point Processes** - This procedure fits statistical models to one dimensional point processes. A one dimensional point process is a process that generates events along a single dimension, usually time or space. The procedure allows for estimation of homogeneous Poisson process models, nonhomogeneous Poisson process models and renewal process models. Tests are also provided to compare multiple samples.

**4. Repairable Systems (Times)** - The Repairable Systems (Times) procedure is designed to analyze data consisting of failure times from systems that can be repaired. It is assumed that when the system fails, it is immediately repaired and placed in service again. Further, it is assumed that the repair time is negligible compared to the time between failures. The goal of the analysis is to develop a model that can be used to estimate failure rates or quantities such as the MTBF (mean time between failures). This procedure differs from the *Distribution Fitting* and *Weibull Analysis* procedures in that it allows for a failure rate that changes as the system ages.

**5. Repairable Systems (Intervals)** - The Repairable Systems (Intervals) procedure is designed to analyze data consisting of failure counts from systems that can be repaired. It is assumed that when the system fails, it is immediately repaired and placed in service again. Further, it is assumed that the repair time is negligible compared to the time between failures. The goal of the analysis is to develop a model that can be used to estimate failure rates or quantities such as the MTBF (Mean Time Between Failures). This procedure differs from the *Life Tables* procedure in that it allows for a failure rate that changes as the system ages.

**6. Dashboard Gage** - The Dashboard Gage procedure creates a gage with a pointer and colored zones.

**7. Frequency Tables** - The Frequency Tables procedure analyzes a single column containing counts. It displays the counts using either a barchart or piechart. Statistical tests may also be performed to determine whether the data conform to a set of multinomial probabilities.

**8. Sampling Distributions** - The Sampling Distributions procedure calculates tail areas and critical values for four common sampling distributions. It also plots the calculated results.

**9. Sequential Sampling** - The Sequential Sampling procedure implements various Sequential Probability Ratio Tests (SPRTs).

Unlike statistical tests that have a fixed sample size, the number of samples required by sequential tests is not predetermined. Instead, after each sample is taken, one of three decisions is made:

- Stop the test and reject the null hypothesis;
- Stop the test and accept the null hypothesis; or
- Continue sampling.

In many cases, the SPRT will come to a decision with fewer samples than would have been required for a fixed size test.

## CHANGES TO EXISTING PROCEDURES

**1. Simple Regression** – The constant term can now be removed from all simple regression models.

**2. Multiple Variable Analysis** – The *Multiple Variable Analysis* procedure now permits drawing box-and whisker plots in the diagonal positions of the *Matrix Plot*.

**3. Oneway ANOVA** – The *Variance Check* pane now conducts F-tests for all pairs of samples, in addition to performing the overall test for equality of variances.

**4. Multiple Regression** – The *Analysis Options* dialog box has been reconfigured to make the available options more accessible.

In addition, stepwise variable selection may now be based on either F-ratios or P-values.

**5. Polynomial Regression** – The *Analysis Options* dialog box has been modified to allow an offset value to be specified.



**6. Distribution Fitting** – Calculated tail areas and critical values may now be saved.

**7. Probability Distributions** – If only one distribution is plotted, users may elect to shade an area of the pdf.

**8. Time Series Analysis and Forecasting** – The time series and forecasting procedures now allow the time indices to be specified in a second column.

**9. Automatic Forecasting** – Several changes have been made:

- The model list has been reworked so that *Random Walk* and *Random Walk with Drift* are separate models.
- The option of estimating or specifying parameters has been structured to be on a model-by-model basis, as is the choice to optimize parameters.
- Additional *Model Selection Criteria* have been added.
- An *Adjustments* button has been added to allow adjustments to be applied before the models are fit.



Statgraphics.Net

C/ Bravo Murillo 350, 1º

28020 Madrid

Tel: 91 571 64 12

Fax: 91 571 21 06

E-mail: [infosg@statgraphics.net](mailto:infosg@statgraphics.net)