



## WHAT'S NEW IN SYSTAT 13

### STATISTICAL FEATURES IN SYSTAT 13

#### *New Features*

#### **ARCH & GARCH Models in Time Series**

Conventional time series analysis procedures assume that the variance of the random (error) terms in the series is constant over time. In practice, however, certain series, especially in the financial domain, exhibit volatility with different levels of variance in different periods. In order to capture and model this phenomenon, *Autoregressive Conditional Heteroskedasticity (ARCH)* models have been developed. Here the variance at each point of the series is modeled using the past disturbances in the series. The ARCH model generally requires a large number of parameters to successfully capture the dynamics of the error variance. The *Generalized Autoregressive Conditional Heteroskedasticity (GARCH)* model, by introducing additional autoregressive terms of the error variance helps achieve parameter parsimony in the modeling.

SYSTAT 13's Time Series Analysis update provides the following:

- Hypothesis tests for ARCH effects: Well-known McLeod and Lagrange Multiplier tests are provided for this purpose.
- Estimation of ARCH and GARCH model parameters by different implementations (BHHH, BFGS, and Newton-Raphson) of the maximum likelihood method with various options for convergence criteria.
- Forecasts for error variances using the parameter estimates.
- The Jarque-Bera test for normality of errors.

#### **Best Subsets Regression**

In the development of a multiple (linear) regression model, it would be nice if the number of predictors in the model developed is small without sacrificing predictive power. The best subsets regression addresses this issue.

- SYSTAT 13 finds best models (choice of predictors) for a given number of predictors, the number varying from one to the total number available in the data set.
- The best model is identified by various criteria such as  $R^2$ , Adjusted  $R^2$ , Mallows's  $C_p$ , MSE, AIC, AICC and BIC.
- SYSTAT 13 then offers to carry out a complete regression analysis on the data set chosen by the user (same as the training set or different) on the best model selected by any of the criteria.

## Confirmatory Factor Analysis

The Factor Analysis feature in SYSTAT 13 now includes Confirmatory Factor Analysis (CFA).

- CFA can be used to test the postulated factor structure based on a priori knowledge about the relationship between the observed (manifest) variables and the latent variables.
- With CFA, SYSTAT allows users to specify the observed variables, a set of latent variables, and their variance-covariance structure.
- A path diagram can be used to specify the hypothesized model.
- SYSTAT 13 estimates the parameters of the CFA model using one of the following estimation options: Maximum likelihood, Generalized least-squares, and Weighted least-squares.
- SYSTAT 13 provides a wide of variety of goodness-of-fit indices to measure the degree of conformity of the postulated factor model to the data. Some of the well-known indices provided are: Goodness-of-Fit Index (GFI), Root Mean Square Residual (RMR), Parsimonious Goodness-of-fit Index (PGFI), AIC, BIC, McDonald's measure of Certainty, and Non-normal Fit Index (NNFI).

## Environment Variables in Basic Statistics

SYSTAT 13 provides environment variables in its Basic Statistics module.

- Environment variables are variables that contain the computed values of various statistics for a given session, a given data file, and given variables.
- These may be directly used in subsequent transformation statements for further processing of the computed statistics.

## Polynomial Regression

SYSTAT 13 provides a direct computation of polynomial regression on a single independent variable. The key features are:

- The order of the polynomial can be up to 8.
- Besides fitting polynomials in standard forms, SYSTAT 13 provides orthogonal polynomial regression.
- SYSTAT 13 reports goodness-of-fit-statistics ( $R^2$  and adjusted  $R^2$ ) and ANOVA with  $p$ -values for all models, starting from the order specified by the user, down to linear (order = 1).
- SYSTAT 13 provides confidence and prediction interval plots along with estimates, and a plot of residuals versus predicted values, as Quick Graphs.

## ***Enhanced Features***

### **Analysis of Variance**

The Analysis of Variance feature now provides:

- Levene's test based on median for testing homogeneity of variances.
- A SUBCAT command that categorizes the desired factors just for the purpose of the analysis.

### **Basic Statistics**

The Basic Statistics updates in SYSTAT 13 consist of

- Standard error and confidence interval for the trimmed mean.
- Winsorized mean, its standard error and confidence interval.
- Sample mode and mode based on kernel density estimate of data.
- Interquartile range

All these statistics are available for columns as well as rows.

### **Bootstrap Analysis**

In most statistical analysis features, SYSTAT has been providing resampling facilities, including bootstrap, whereby the same analysis is carried out on each of the samples and complete outputs produced for each of them. SYSTAT has recently been providing, in a few features, analysis of the bootstrap outputs, summarizing the key parts of the outputs by histograms, computing various types of bootstrap estimates, their biases, standard errors, confidence intervals, and  $p$ -values. In SYSTAT 13, this facility is added to Hypothesis Testing and enhanced in Least-Squares Regression.

- In Hypothesis Testing, SYSTAT 13 provides bootstrap-based  $p$ -values and histograms of the test statistic obtained from the bootstrap samples. Apart from the usual  $p$ -values of the tests, users can now request bootstrap-based  $p$ -values in all tests for mean (one-sample  $z$ , one-sample  $t$ , two-sample  $z$ , two-sample  $t$ , paired  $t$ , Poisson) and variance (single variance, two variances and several variances)
- In Least-Squares Linear Regression, a choice of bootstrapping residuals has been included. Bootstrap estimates of the regression coefficients, their biases, standard errors, and confidence intervals are then computed based on these.

### **Crosstabulation**

Updates provided in the Crosstabulation (XTAB) feature include the following:

- Relative Risk: In a 2 x 2 table, the relative risk is the ratio of the proportions of cases having a 'positive' outcome in the two groups defined by row or column. Relative Risk is a common measure of association for dichotomous variables.
- Mode: SYSTAT 13 gives an option to list only the first N categories in a one-way table (frequency distribution). This is done by adding a MODE = N option to the PLENGTH command within XTAB.
- Saved results with:
  - all requested columns in Multiway: Standardize
  - value labels of the input variables for the corresponding columns of the saved results file.
- Output categorized appropriately based on the type of table, and reorganized table of measures.

## Fitting Distributions

SYSTAT 13 performs the estimation of parameters using the maximum likelihood method for the following distributions:

- Beta
- Chi-Square
- Erlang
- Gamma
- Gompertz
- Gumbel
- Logistic
- Log-Logistic
- Negative Binomial
- Weibull
- Zipf

## Hypothesis Testing

- The Hypothesis Testing feature has been strengthened by adding tests for mean vectors of multivariate data.
  - One-sample Hotelling's  $T^2$  test for mean vector of multivariate data equal to a known vector.
  - Two-sample Hotelling's  $T^2$  test for equality of two mean vectors of multivariate data.
- For two-sample z, two-sample t, and test for two variances, users can now directly input data in a layout where the data across the samples appear in different columns. This is in addition to the current indexed layout.

## Least Squares Regression



SYSTAT 13 allows you to save standard errors and confidence intervals obtained from Least Squares Regression.

## Logistic Regression

SYSTAT 13 offers more intuitive ways of analyzing binary, multinomial, conditional, and discrete choice models. Specifically:

- Simplified and easy user interface and command line structure to analyze binary, multinomial, conditional, and discrete choice models separately.
- Option to specify the reference level for the binary and multinomial response models.
- Simpler form of input data to analyze matched sample case-control studies with one case and any number of controls per set.
- For Discrete choice models, SYSTAT 13 provides two data layouts: 'Choice set' and 'By choice' to model an individual's choices in response to the characteristics of the choices.
  - In the 'Choice set' layout, choice set names for groups of variables can be defined, and variables can be created, edited, or deleted.
  - In the 'By choice' framework, the choices sets already defined can be used in the data for the analyses.

## Nonparametric Tests

Updates in Nonparametric tests are:

- Jonckheere-Terpstra test as an alternative to Mann-Whitney test: The test is used when the treatments are ordered in terms of the response. This test is based on the sum of the  $k(k-1)/2$  Mann-Whitney counts (for  $k$  treatments).
- Fligner-Wolfe test as an alternative to Mann-Whitney test: The test is used when one of the treatments acts as a control, to test equality of response to control vis-à-vis all other treatments, with a one-sided alternative. This is a Mann-Whitney test with two groups when the control is one group and all other treatments together form another.
- Post hoc (multiple comparisons) tests of Dwass-Steel-Critchlow-Fligner and of Conover-Inman which can be used as a follow-up when a Kruskal-Wallis test shows significance.
  - The Dwass-Steel-Critchlow-Fligner test is based on an asymptotic Studentized range distribution
  - The Conover-Inman test is based on asymptotic Student's t-distribution.
- A multiple comparison test due to Conover as a follow-up of Friedman test showing significance.

## NEW GRAPHICS FEATURES IN SYSTAT 13



- Colors in terms of their RGB (Red-Green-Blue) component values. This is available for specifying the color of elements, axes and frame colors.
- Gradient style for surfaces through the dialog box interface.
- Labeling dots in dot (summary) charts.
- 45 built-in colors as against the 12 available in previous versions.
- Colors for overlaid graphs, pie and stacked charts in such a way as to provide more contrast between adjacent elements.
- Stacked bar charts with grouping variables where a stacked chart is drawn for each group, and all the charts are laid out in the same frame.
- Individual border displays on plots allowing you to suppress the display along any given border, or specify different kinds of border displays along the two borders in all two-dimensional plots.
- Separate multiple slices in a pie chart.
- Numeric case labels in plots, multivariate displays and maps using numeric variables.

### **OTHER NEW FEATURES IN SYSTAT 13**

- Autohide the Workspace and Commandspace.
- Define choices using the Choice Token dialog box where each choice may be linked to a SYSTAT command script.
- Data Edit Bar which allows you to navigate to any cell in the Data Editor, and view/edit data values.
- File Comments button to access data file information by a simple click of the mouse.
- Default Format for Saving Command Files.
- Drag-and-drop data into SYSTAT's Data Editor from editors that support dragging of content.
- Embedded Format Bar, Data Edit Bar and the Graph Editing toolbar in the respective tabs of the Viewspace.
- Open and execute legacy command files by inserting a VERSION command at the top of the file.
- View Toolbars using the View menu.
- Windows XP Style Grids for SYSTAT's Data/Variable Editor, Data: Transform: If Then Let dialog box, and Data: Select Cases dialog box.
- Trim leading and trailing spaces in string data.
- Close data files using the context menu of the Data Editor, or the CLOSE command.
- Default variable format for newly created numeric variables in the Data Editor, independent of the numeric output format.
- Save view mode data files.
- Import business objects where you may use a "Business Objects Universe" as a data source.
- Locales and digit grouping settings to be used while displaying numbers in the Output Editor.
- NODE command to set and suppress Output Organizer node and collapsible link captions.
- ACTIVE Command to activate a file that is in the view mode.
- Additional built-in functions
- FOCUS command to switch focus to the Data Editor, Graph Editor or Output Editor.
- Define and call macros in your command scripts.
- Autocomplete command arguments, options and option values.



- Optional command keyword coloring in the Commandspace.
- Tabbed dialog boxes with the tabs arranged vertically allowing more tabs to be easily accessible with just a single click of the mouse.
- Attempt to restore a session through the Rescue Report dialog box.
- Additional shortcut keys.
- Enhanced Status Bar that allows you to set the page width, and toggle the states of the Insert, Caps Lock, Num Lock and Scroll Lock keys on the keyboard.
- Download Themes dialog box wherein you may choose which themes to install.
- Versioning of theme files so that you will have the option to upgrade your theme file whenever a newer version is available on the SYSTAT server.
- Prompt to save the current theme when you apply a theme.
- Copy a cell in the Data/Variable Editors and paste it into a column.
- Open multiple data files with output ordered chronologically or based on the input data file.
- ELSE option while recoding variables which will allow you to recode all values other than a given set of values to a certain specified value.
- Store and retrieve current settings for the active data file, value label display format, and variable label display format.
- Unlimited page width using the PAGE NONE command.