TECHVIEW
SOFTWARE

Statistics and More Statistics

Allen Rawitch

SYSTAT 8.0 is the latest release of a statistics and graphing program for Windows-based computers. It has evolved over several versions to become sophisticated, yet user friendly. The program is versatile and permits a broad variety of statistical analysis and estimates to be performed on data that has been entered directly or imported via a number of common formats. An extensive set of statistical functions is available in the system (Table 1). A graphical interface allows most operations to be carried out by point-and-click input. In most cases, when analysis routines are executed, the program output can be data tables or basic plots. Users may select specific plots or graphics to suit their needs. Output options include several graphs with three-dimensional formats.

SYSTAT 8.0
SPSS Inc.
Chicago, IL.
$995.00,
$695.00 (academic).
800-521-1337
www.spss.com

In addition to the wide breadth of statistical analyses listed in Table 1, the programming features of SYSTAT permit the user to customize the program extensively. SYSTAT has a well-documented command language, which allows one to write command scripts or batch files that can be saved for repeated or frequent procedures. In addition, BASIC programs may be written to further expand the program's capabilities. The combined graphic and analytical functions of SYSTAT serve the needs of a wide variety of researchers, but the programming features make it more suitable for sophisticated statisticians than for the casual or novice user.

Data is entered into SYSTAT 8.0 by means of a fairly standard table or matrix format, with columns and rows identified by variable names. These can be formatted as either numeric or text fields. Data may be imported into the program from Excel, Lotus, dBASE, or ASCII file formats and can be exported in the same formats and DIF (Data Interchange Format), as well. An additional feature of the program is its ability to use functions from, and share data or output with, the BMDP program (Meridian Marketing Group).

The author is in the Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, KS 66160, USA. E-mail: arawitch@kumc.edu

The opening or main window of SYSTAT 8.0 is divided into three panes, one of which is active at all times. Switching between the panes is accomplished by clicking in the appropriate window. The output organizer pane of the main window provides a tree-like format reflecting the original data file and any output of analysis or graphing derived from it as subelements organized in a tree structure. Users can cut, copy, paste, or delete the individual subelements (tables, graphs, and the like) in order to modify the information displayed in the output panel. The output organizer also allows the user to quickly and directly move to specific portions of the output. The second pane in the main window contains the output pane, which displays results of any operation, including file input/output or, more commonly, results of statistical analyses. The third pane is a command pane, which, when selected, allows the entry of executable commands, the running of batch files, and the display of a command log associated with the operations. Contents of the command pane may be saved and can be run later as a batch file for repetitive, customized calculations. The organization of the main window is easily reconfigured with standard Windows 95 procedures.

The second window used by SYSTAT is a data window. It is arranged in rows and columns representing cases and variables, respectively. The data window is used for both initial data input and modification. Rows, columns, and blocks may be cut, copied, and pasted within it. Variables are defined in a dialog box that opens when the variable name segment (the top cell in the data matrix column) is double-clicked. Variable names can be assigned there along with the definition of the variable format or data type. Data transformations (application of a mathematical function or a variety of statistical functions) may be performed in the data window as well. In addition, subsets of data may be highlighted in the data matrix with the mouse and used as the basis for new data files or calculations.

The third and last basic window type in SYSTAT 8.0 is the graph window. Although graphs can be initiated from a menu bar in the main application screen, once a graph is generated, its properties may be modified in the graph window. Text may be inserted, fonts changed, line color and attributes changed, symbols changed, and legends added, as well as specific points labeled. Functions in this window allow the user to rotate three-dimensional graphs, tune tension settings for smoothers, and select a set of cases from the data matrices. Graphs may also be exported or printed from this window.

Dialog boxes are available for most of the menu options in all three of these windows, and window format and organization can be modified in each case.

In addition to the common forms of bar, dot, line, profile, pyramid, and pie charts in two- and three-dimensional formats, SYSTAT 8.0 creates density charts, quantile and probability plots, scatter plots, multivariate display plots, function plots, maps (in various perspectives), and quality-control charts. Finished graphs may be printed directly from the graph window or saved in standard graphic formats, such as Windows bitmap (BMP), JPEG, or two different Windows metafile formats. Statistical analyses and other results are saved as SYSTAT data files with the extension ".syd." The ability to edit and format graphs and charts and to change perspective or view of the graphics makes SYSTAT very useful for producing publication-quality figures. Symbols, labels, line attributes, axes, and scale can be edited after the initial graphic output is obtained. Plots may be overlaid and multiple graphs placed together in a straightforward manner.

One of the strengths of the SYSTAT 8.0 package is a set of extensive manuals that are well written and illustrated. In addition to the "Getting Started" booklet, which contains orientation examples and exercises to allow the user to begin working reasonably quickly, there are five additional sources of support.

1) A language reference book addresses global commands used by SYSTAT and provides a discussion of BASIC programming and specific functions used in graphic, statistical analysis, and data manipulation.

2) A data manual describes data entry procedures and import/export strategies among common file formats. It addresses the manipulation and transformation of data, as well as variable manipulation and grouping. This manual also discusses matrix manipulations and matrix algebra. Virtually every operation in the manual is illustrated with an example.

3) The well-illustrated graphics manual provides an excellent overview of the graphics and tables that can be created with the program and how they are obtained.

4) A statistics manual, containing more than 1000 pages, combines a well-organized discussion of statistical approaches and treatments, with illustrations showing how they can be implemented. These include the application and limitations of bootstrap approaches to data evaluation. The manual contains examples of SYSTAT scripts, programs, and forms of output.

5) Users can subscribe to a SYSTAT listserv, sponsored by the publisher, to receive online support for the program. Additional online help can also be obtained from the "Knowledge Base," a series of frequently asked questions and their answers. As with the printed documentation, the online support is excellent.

All of the manuals contain extensive references for the models and analyses they describe. One can retrieve examples of procedures and analyses, and the scripts used to execute them. These scripts can be cut and pasted to meet a user's specific needs. A good discussion of the SYSTAT executable command structures and their application is presented in the "Getting Started" manual for the system. A very nice feature of the program is its application gallery of data and analyses relating to specific research fields. These include research areas such as anthropology, astronomy, biology, chemistry, engineering, environmental sciences, manufacturing, medical research, psychology, sociology, statistics, and toxicology.

Despite its complexity, SYSTAT is a very stable, well-behaved program. No crashes were encountered during this evaluation. When analyses were attempted with inappropriate data or parameter identification, the program returned an error message specifying the nature of the deficiency, or inappropriate parameter identification, and allowed backtracking to fix the problem. Although it is certainly likely that under some circumstances, a crash or lock-up may occur with SYSTAT, it is clear that this is not a significant problem.

In spite of the fact that SYSTAT 8.0 is a very sophisticated software package targeted to the serious user of statistics, its user interface is well done and could allow even novices to use its basic functions, as long as they were aware of the appropriate statistical treatment required for their data. The programmers have thoughtfully included many datasets and examples from a number of research disciplines, along with examples of applications that use these datasets. The availability of these datasets eases the learning curve for first-time users, enabling them to concentrate on functions without having to enter large amounts of data.

SYSTAT 8.0 is a very powerful and flexible data analysis program with strengths that include extensive written and online documentation and support, facilitated data entry, and extensive and flexible graphical output. The menu-driven functions of SYSTAT 8.0 are sufficient to address most users' requirements for the statistical analysis of data. The ability to create executable BASIC programs and scripts, which may be run as batch files, is useful for the more sophisticated statistics user. The program should serve the needs of a wide variety of researchers. SYSTAT's depth of function comes with a price, however—users should expect to spend time learning to use all of the functions of the program. This task is made easier, fortunately, by numerous examples and excellent, well-written documentation. An afternoon session spent with the online examples and the data and documentation in the "Getting Started" manual should provide the average user with the background necessary to master the basic features of SYSTAT 8.0. That such a complex application as SYSTAT can be so easily mastered is probably the program's greatest strength.

SYSTAT 8.0 is provided on a single CD-ROM (3.5-inch floppy disks available by arrangement with SSPS). Its installation is smooth and without remarkable complications. This release requires either Windows 95 or Windows NT 4.0 as an operating system. Hardware requirements include at least an 80486 CPU with a floating point processor and 16 MB of RAM. At least 32 MB of available hard disk space is required for proper functioning of the program. The user is prompted during the installation to enter a 38-digit license validation code. I evaluated SYSTAT on two different Pentium-style computers running Windows 95. The high-end computer was a 300-MHz machine with 64 MB of RAM and the other was a 90-MHz system with 32 MB of RAM. The software performed well in both environments.

STATISTICAL FUNCTIONS AVAILABLE WITH SYSTAT 8.0

Function	Description
Descriptive statistics	Basic statistics, including mean, min, max, *n*, and standard deviation
Stem and leaf plots	Calculation of Cronback statistics
Crosstab analysis	One-way, two-way, and multiway
t-Test analysis	One sample, paired, and two-group calculations
Analysis of variance	Both estimates and hypothesis testing
A design of experiment feature	Generates design matrices for multifactorial and multilevel experiments for application to quality control and product improvement
Nonparametric testing	Including Kruskal-Wallace, Wilcoxon, Friedman, and Wald-Wolfowitz approaches
Correlations	Simple, set, and canonical
Regression analyses	Including linear, logit, probit, nonlinear and two-stage least squares analyses
Time series analysis	12 different calculation functions
General linear and loglinear models	Both estimate and hypothesis testing
Classification analysis	Including discriminant analysis and several clustering analysis tools
Spatial statistics	Calculations of a variety of statistics for two- and three-dimensionally oriented data sets
Survival statistics	Explore grouped, right, and interval censored data and estimate models
Classification and regression tree calculations	Produce graphic representations of data distribution and goodness of fit for systems calculations with catgorical or continuous dependent variables
Test item analysis	Both classical and logistic
Data analysis to aid in specifying signal detection parameters	Facilitate analyses employing signal detection theory for analysis of weak signal or response superimposed on noisy background
Path analysis and structural equation modeling	Apply a Reticular Action or Near Approximation model to fit path analysis models to correlation matrices
Conjoint analyses	Fitting of metric and nonmetric models to observed data