A Carnival of Stats

Bruce McCallum

tatistica 5.1 for Windows is a powerful tool for scientific computation, epidemiological studies, and industrial systems analysis. The program has multimedia presentation functions and contains a help function in the form of an online Stats Advisor. Use of the advanced techniques in the program, however, requires extensive familiarity with the five-volume instruction manual.

Data Management and User Interface

First-time users will find data management made easier than previous versions, with intuitive dialogs and menu buttons. Data entry follows the conventions found in spreadsheet programs, rather than the variables (columns) and cases (rows) format of standard statistical packages. *Statistica* spreadsheets support drag-and-drop moving and copying, in-cell editing, split-screen scrolling for long columns of data, automatic column width adjustment, pop-up flying menus, and automatic extrapolation of values in blocks of data. Most of these features are also found in *Statistica* scrollsheets, a spreadsheet-like output window for displaying results of statistical analyses. Data entered in SPSS, SAS, Excel, Lotus, Quattro formats, and dBase Oracle, Sybase, and Access database files can be imported into *Statistica* without first saving them as ASCII text, if the proper filters are installed during setup. A basic test was performed on a large Excel 97 data file with good results; *Statistica* stripped off underlying formulas and recognized specially formatted numbers, such as dates, without difficulty. In combination with a data management module called Megafile Manager, *Statistica* can handle 32,000 variables and unlimited cases of up to 255 characters per cell, depending on the amount of installed RAM. Only one data spreadsheet can be open at any time in each module.

The author is in the Department of Anesthesiology, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA. E-mail: mccallum@mcw.edu

> **Statistica 5.1 for Windows**
> StatSoft, Inc.
> 2300 East 14th Street,
> Tulsa, OK 74104, USA.
> $995 U.S.;
> $1245 overseas.
> 918-749-1119
> www.statsoft.com

Alas, the program still relies on restrictive formats for some data management chores. For example, imported data defaults to an eight-character text string or whole decimal numbers. Making global changes to columns, such as recalculating or reformatting variables, requires the use of another variable specification spreadsheet window. Changes made to variables from the specification window cannot be undone. These requirements prolong data management tasks, particularly for entering decimals or names. One advantage of the approach, however, is that *Statistica* automatically matches text values with underlying numerical values so that distributions or frequencies of text variables can be analyzed.

The program interface includes a variety of Windows-specific task bars, dockable windows, menu icons, text output windows, and floating automation buttons. Users are presented with a task bar at startup for switching between a group of related procedures called modules. Modules have access to spreadsheet and graphical functions, but each module contains a limited set of statistical analyses. With multitask-

A summary of some of the many display options in *Statistica*.

ing, each module independently processes the same data set at the same time. For example, descriptive statistics on one set of data can be examined in the Basic Statistics module, while the same set of data is simultaneously analyzed in the ANOVA/MANOVA (analysis of variance/multivariate analysis of variance) module. Changes made to data in one module can be merged with changes in another module or saved separately. Customizable toolbar buttons that control operations in document windows

provide pop-up tips to help users perform difficult tasks. In addition, floating buttons can automate repetitive tasks, like saving or merging data, via STATISTICA BASIC, a program-specific command line language.

After data have been chopped and diced by *Statistica*'s statistical procedures, a queue of document windows presents the next possibilities. One option for saving or printing results is the text and graph output window, which exports in rich text format (RTF) to word processors. Accumulated results from analyses and graphs are stored in a separate, continually updated document window. Setting the program to direct text or graphics to the window can be frustrating, however. Sometimes, the program freezes or cannot locate data from previous work. A safer option is the Workbook manager, which saves and links every document window in an active session to a data set. When the Workbook manager reopens the document, it opens the appropriate data set simultaneously.

Statistica offers full OLE (object linking and embedding) compatibility with other Windows applications and files. Graphs or pictures from Windows applications can be embedded in *Statistica* graph windows, or *Statistica* graphs can be linked to Windows programs like Microsoft PowerPoint and updated from within *Statistica*.
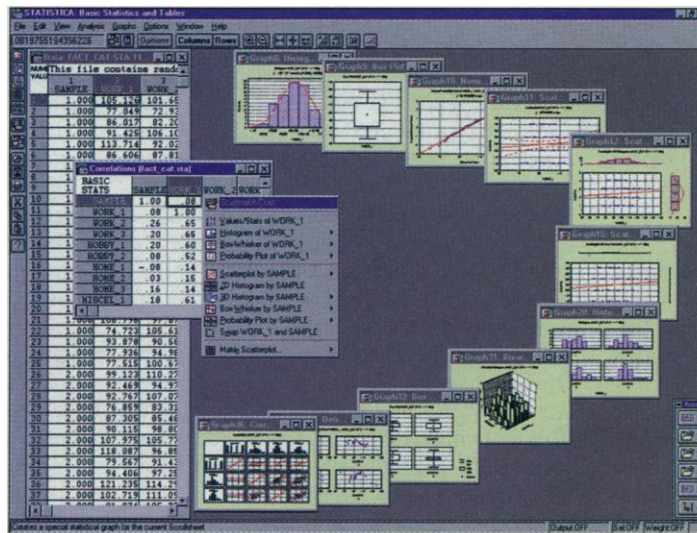
Statistical Analyses

Statistica has two approaches to statistical analyses. Quick Basic Stats furnishes instant access to correlations, frequencies, or descriptives for a block of data or highlighted variables from any data window. A handy little Probability Calculator allows users to interactively explore normal and non-normal distributions, such as t, F, Z, or Chisquare, by adjusting parameters. Output generated by Quick Stats can be reversed using up to 16 "undo"s. Exploratory data analysis or hypothesis-testing of main effects follows another path. The Module Switcher provides access to grouped suites of statistical analyses. A startup menu offers individual tests available in each module, with one panel for creating models and another for testing results. Graphical "continue" buttons lead back from results windows to the results panel so that the user is not lost in a sea of stats.

The Basic Statistics and Tables module offers more extensive correlation, frequency and descriptive tests, as well as independent and dependent t-tests and one-way ANOVA. A potential weakness of predesigned statis-

tics modules is that users can obtain significant results without examining the strength of relationships (equality of variation) or the reliability of sample populations (normal distribution). *Statistica* addresses sample-size problems by providing graphical tests of normal distribution via histograms or box-and-whisker plots, directly from the results panel. Diagnostic tests for the equality of variation include the *F* test or the more robust Levene's test. Violations of normalcy indicate the need to switch to the Nonparametrics and Distributions module, while violations of the equality of variation assumption may suggest further analysis of variation in the ANOVA/MANOVA module.

The ANOVA/MANOVA module handles models for comparison of means, complex designs for interactions between groups and within groups, incomplete nested designs, and analysis of covariance and multivariate designs. Anyone whose research involves extensive use of repeated measures designs will be pleased to discover that *Statistica* offers both univariate and multivariate contrasts of means between groups and within groups. To verify the independence of these interactions, *Statistica* automatically runs tests of symmetry by using Greenhouse-Geisser correction and Huynh-Feldt correction, as well as tests of sphericity with Wilks' Lambda, Rao R, Pillai-Bartlett Trace, and V. Post-hoc comparisons of means for factorial designs include LSD, Scheffé, Duncan, Newman-Keuls, and Tukey. The startup panel for variable selection and model design is easy to use and extremely flexible. The results panel offers complete control over the interpretation of relations between variables, in addition to prohibiting analyses that violate ANOVA assumptions. *Statistica* automatically handles missing data designs by pooling results, rather than eliminating cases. For more complex patterns of missing data, nested models are available. Overall, the ANOVA/MANOVA module is the most thorough analysis of its type in the program or in other programs known to this reviewer.

Fitting data with the Nonlinear Estimation module was not satisfactory. The only growth and decay functions available were logistic, probit, exponential, and piecewise linear regression. Additional functions must be entered by the user—for example, entry of a dose-response function of the form where b_0 is the maximal response, b_1 is the slope factor, and b_2 is the half-maximal response. *Statistica* fits experimental data well by using a quasi-Newton algorithm and plotted fitted values on a graph with the fitting function displayed, but no upper and lower confidence levels were available. Researchers who plot Boltzmann probabilities, Hill equations, or other pharmacokinetic functions will need another program for those tasks.

Survival Analysis is available in another module with options for categorized Life Tables or continuous survival Kaplan-Meier analysis. *Statistica* handles subjects with complete or partial outcome information. The program also estimates the probability of survival with nonparametric tests and assesses significant differences between survival rates. Health scientists, social scientists, and industrial product designers will find this module very useful. Structural Equation Modelling is a unique statistical module in *Statistica* that theoreticians modeling pathways in a complex set of interactions may find helpful.

Graphics
Statistica consistently garners superior ratings for its graphics. The main strengths of the program's graphics are quick accessibility and the powerful customization tools available for each graph. A highlighted block of data can be converted into a graph with a simple button click (see figure). Results menus containing standardized graph buttons can instantly transform statistical analyses into visual results with superimposed confidence levels or fitting functions. More-complex graphs are available in a menu-driven panel for selecting variables and chart types, and customized chart templates can be saved for future use. After plotting, graph properties can be modified in a variety of ways. Multiple graphs can be merged or pasted into blank graphs. Graphs can be embedded in other applications and dynamically updated thanks to OLE implementation. *Statistica*'s features include a brushing tool to allow users to temporarily remove outliers (points outside of a particular area), in order to explore their effect on analysis. The program can animate entire correlation matrices to explore the effect of one relationship on different strata or highlight subgroups in scatter plots. If an output window has a matrix format or a format where a cursor position may indicate more than one variable, then *Statistica* automatically selects predefined bivariate graphs for the specified pair of variables.

Documentation
A full installation of *Statistica* comes with 200 MB of multimedia animated overviews providing quick introductions to most topics in the program. Inside *Statistica*, an online electronic Statistical Advisor guides users through a series of questions and answers. Information from the online help is extensive enough that a knowledgeable statistician might possibly navigate the program without consulting manuals. Use of advanced techniques will require the manuals, however. Documentation is extensive; five volumes totaling more than 5000 pages ship with the program. Fortunately, instructions

are clearly written for beginners as well as adept analysts. Every chapter contains a section on overall theory, a set of examples, and technical discussions of features and statistical functions. An introductory option is the Electronic Statistics Textbook in HTML, available as a free, 5-MB download from the company Web site. Technical support from the company is very responsive, both online and over the phone. The company also hosts seminars in major cities in the United States and worldwide.

In conclusion, *Statistica* maintains its top position among standardized statistical computer programs in today's market. The program's rich graphing features and structured statistical modules make separate programs redundant or unnecessary. It should be noted, however, that the full version of *Statistica* lacks a module for power analysis, a procedure increasingly required by editors of clinical journals. The module, which has recently been released for $495, can be used as a stand-alone product, or in conjunction with *Statistica*. One alternative to the full version is a "lighter" version of the program containing only basic hypothesis testing modules, available for about half the price.

A reasonable question to ask is whether one should pay almost $1000 for *Statistica*'s functions, which are also available on less elegant, but less expensive, programs. The answers lie in the program's performance. *Statistica*'s multiple tests of assumptions and associated graphical exploratory tools make interpretation of data easier and more reliable. Authors cite *Statistica* in the methods section of their papers because of the program's high standards, commitment to accuracy, and technical sophistication for output. The documentation provides a brief education in statistics and will be welcomed by researchers who have had a course in statistics. They can readily familiarize themselves with *Statistica*'s basic concepts and then advance to more specialized analyses as the need arises.

An Intel 386 processor or greater—running Windows 3.1, 95/98, or NT—with 304 MB of hard disk space and 4 MB RAM is required in order to operate *Statistica*. The program was tested on a Pentium 100-MHz processor with Windows 95. The program ships on a CD-ROM, but floppy disks are available. International versions are available in Japanese, German, French, Polish, Russian, and Italian. A commercial update is due in the second quarter of 1999, and a new release is scheduled for late 1999. The program is year 2000 compliant. A version for the Macintosh is available commercially, but it is out of date and no new versions are planned.