Review of STATISTICA 6

John Newell
National University of Ireland, Galway

john.newell@nuigalway.ie

The first (DOS) version of STATISTICA was released in March 1991 by Statsoft. The company claim that the release of STATISTICA 6 marks a new generation of statistical software, offering "unique functionality and usability features that no other competing product can offer". In particular they mention the software's user-friendly interface, virtually endless possibilities for customisation, enhanced options for Web-enablement, and claim that STATISTICA 6 is the only data analysis software on the market that features built-in Visual Basic scripting (ie the same as Microsoft Excel or Microsoft Word).

Their claim seems well founded, however the software is pricey (over £3, 000 for a single stand-alone copy STATISTICA-BASE plus the add on modules although academic discounts are advertised). A reduced feature student version is available also although there was no price for this product or the price for an academic site licence available at the time of this review.

The STATISTICA Family of Products

The core unit of STATISTICA 6 is "STATISTICA-BASE", a stand-alone package that includes descriptive statistics (e.g. correlations, t-tests and other tests of group differences, frequency tables and crosstabulations), multiple regression methods, nonparametric methods, ANOVA/MANOVA routines, distribution fitting modules and all of STATISTICA's graphics tools.

In addition to STATISTICA-BASE, the following 8 add-on modules are available: Advanced Linear/Non-Linear Models, Multivariate Exploratory Techniques, Quality Control Charts, Process Analysis, Design of Experiments, Power Analysis, Neural Networks.

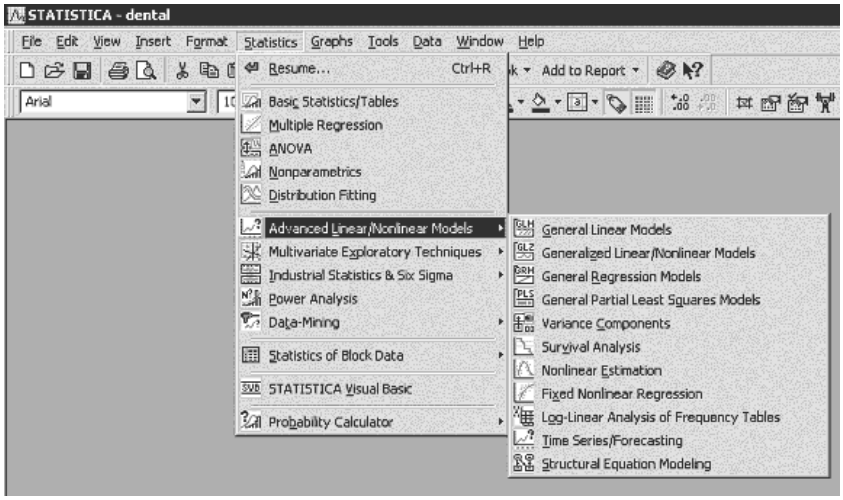

Figure 1

Installation and System Requirements

The software shipped on a single CD and included three manuals; a getting started manual entitled "The Small Book" (150 pages), a quick guide to Visual Basic scripting (150 pages) and a system reference (1100 pages). The CD also included an interactive guide to the product and electronic manuals.

Supplier's contact details

StatSoft Ltd
21-23 Mill Street
Bedford, MK40 3EX

T: 01234 341226
F: 01234 341622
E: info@statsoft.co.uk
W: http://www.statsoft.com/uk/welcome.html

STATISTICA 6 is compatible with Windows 95/98/NT/2000/XP/Me with minimum requirements of a Windows-compatible CPU, 32MB RAM and 70MB of hard disk space to install STATISTICA-BASE plus the 8 add-on modules. This review is based on STATISTICA-BASE plus all add-on modules running under Windows 2000 on a Dell Precision 420 dual-processor Pentium 650 with 256MB Ram.

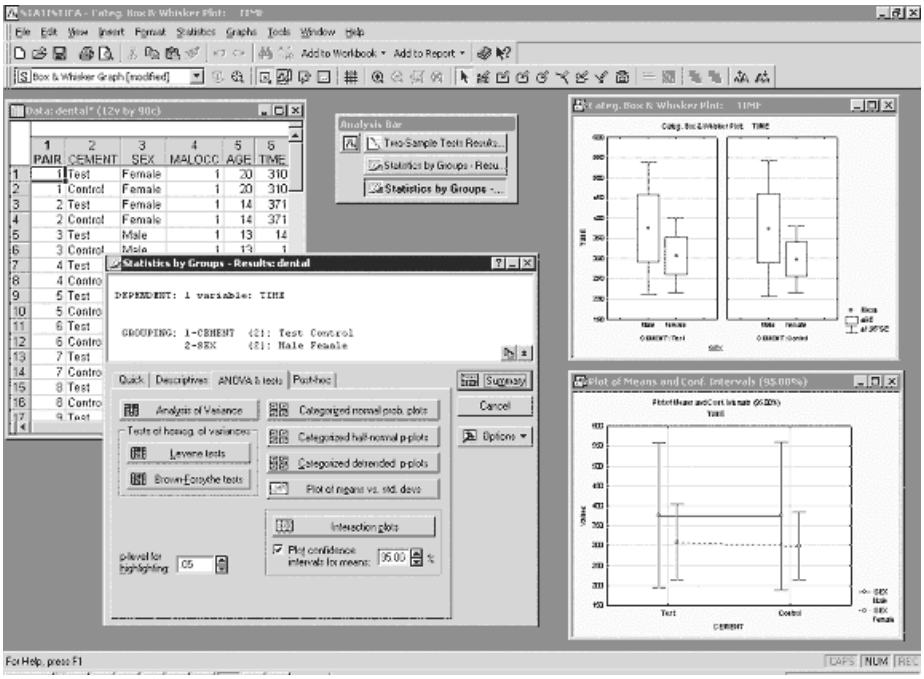

Figure 2

Interface

In previous versions of STATISTICA, each module (e.g. Basic Statistics, Data Management etc) ran in a separate window while in the current version all of the modules are available in a single drop down menu (Figure 1). This is an immediate improvement as multiple analyses can now be open simultaneously in the same window, and can be performed on the same or a different dataset (multiple datasets can be opened simultaneously).

All current "analyses" (eg a graph, an ANOVA, a Multiple Regression etc) are represented by buttons on the "Analysis Bar" (Figure 2) allowing the user to quickly return and continue (or modify) an analysis (e.g. change variables, edit graph features or specify additional tests). This is a useful and time saving feature.

Menu customisation is extremely easy where modules can be moved into new toolbars in order to avoid clutter. For example, toolbars could be created to contain the type of analyses typically needed in a particular statistics course (Figure 3).

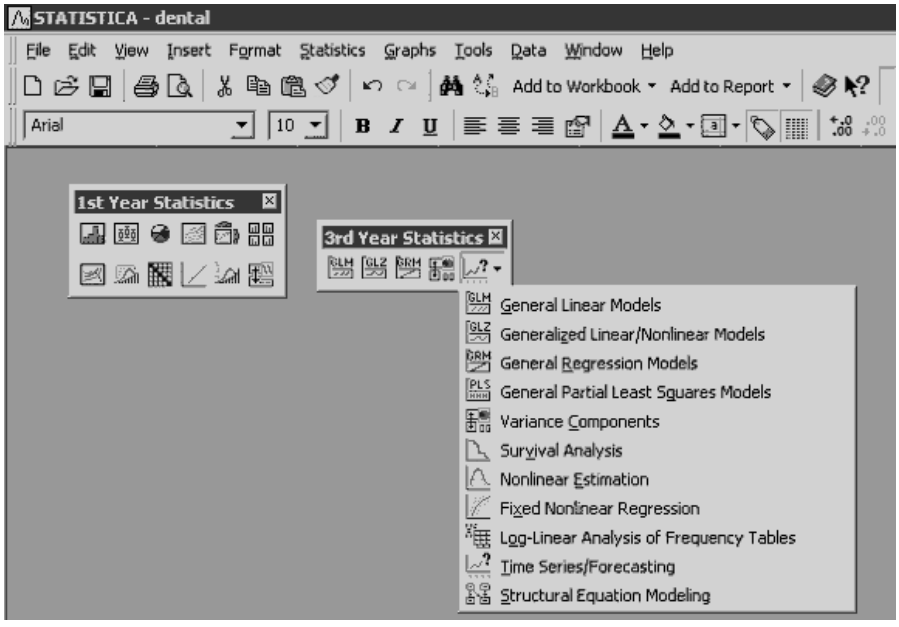

Figure 3

Once an analysis is selected, there are three alternative user-interfaces (Figure 4): Quick-specs dialogs (which prompt the user to specify the necessary variables), Wizards (which guides the user through the selection process) and Analysis Syntax (writing command syntax). Many of the design based modules offer a choice of either the traditional SAS® language or STATISTICA's VGLM language (Figure 5). STATISTICA will automatically generate the complete set of syntax statements for any design specified via the Quick-specs dialogs or the Wizard allowing a log of complex and customized designs to be saved for future use on new datasets.

Data Handling

STATISTICA supports importation of various text formats in addition to data generated by Dbase, Excel, Lotus and SPSS. Data are contained in a spreadsheet and a variety of spreadsheet data management features are available such as sorting, recoding, standardising etc and drag and drop facilities are available for direct data manipulation. There are virtually no limitations in terms of the number of cases, variables, text label length and case names. The size of the spreadsheet, the number of variables, and the text labels appear to be limited only by the available hardware.

Statistical Features

There are an impressive amount of analysis modules available in the Statistics drop down menu, which are not apparent at first glance (Figure 1). The General Linear Model analysis option for example offers 16 types of analysis (see Figures 1 and 4) and STATISTICA has several statistical features that commonly appear in stand-lone applications (e.g. Power Calculations, Tree Based methods etc).

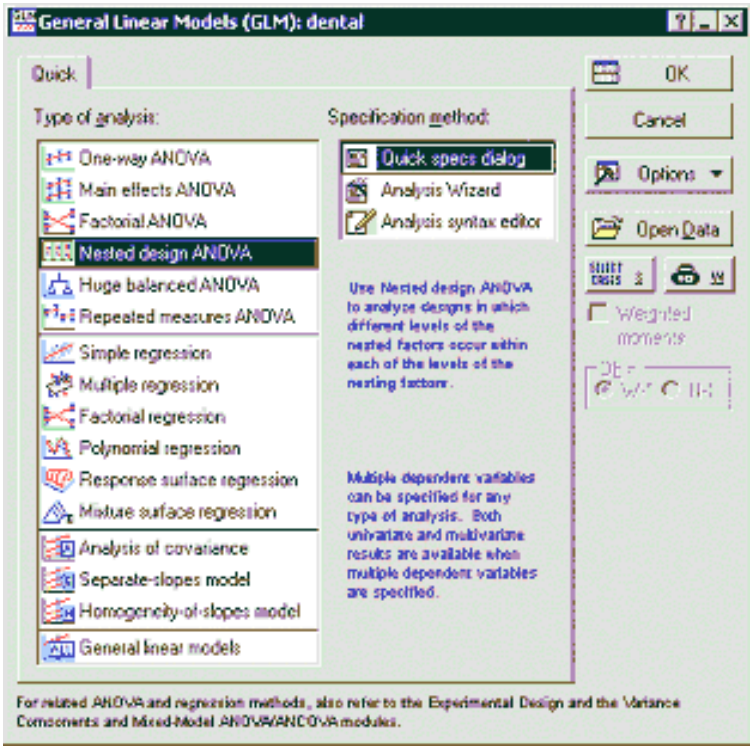

Figure 4

As this review's emphasis is on using STATISTICA for teaching purposes, suffice to say STATISTICA has all of the necessary modules needed for statistics courses ranging from general introductory courses to courses common in a 4 year honours statistics programme such as experimental design, multivariate analysis, generalised linear models, survival analysis, logistic regression etc. Statistics courses for engineers are well catered for also as there are extensive modules available for statistical process control and industrial statistics. Additional statistical modules for a Psychology degree, such as Factor Analysis, Principal Components and Classification Analysis, Correspondence Analysis and Multidimensional Scaling are present also.

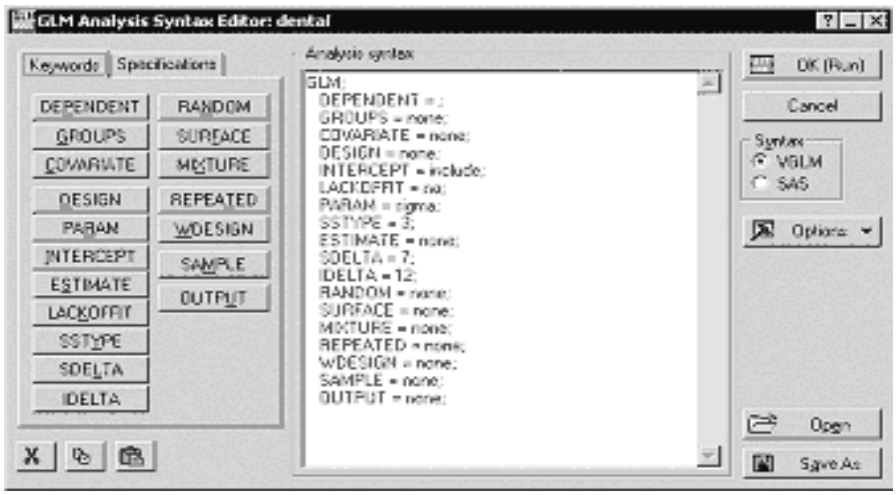

Figure 5

In addition to many of the modules common in most commercial statistical packages, STATISTICA 6 includes a suite of five new modules for analysing linear and nonlinear models:

1. General Linear Models (including algorithms for fitting overparameterized and sigma restricted models, Type I through IV sums of squares for unbalanced and incomplete designs and effective hypothesis decomposition),

2. General Regression Models (incorporating stepwise and best subset methods to build models for highly complex designs, including designs with effects for categorical predictor variables extended to analyses of problems with multiple dependent variables),
3. General Discriminant Analysis Models (allowing the user to specify simple and complex ANOVA and ANCOVA-like designs),
4. Generalized Linear Models (for investigating linear and nonlinear relationships between a continuous, or binomial, multinomial, or ordinal multinomial categorical response variable and categorical or continuous predictor variables. In addition, routines for Generalized Additive Models are included),
5. General Partial Least Squares Models (offers a selection of algorithms for univariate and multivariate partial least squares problems).

All of the above routines offer stepwise and best subset options, crossvalidation, extensive residual analyses and plots for both the training and cross-validation samples and of variance components. As such, any of the modules offered in STATISTICA-BASE (e.g. ANOVA, regression etc) can be considered as subsets of those offered through the General linear model modules.

In theory modules for computational projects requiring computational inference such as bootstrapping, randomisation and permutation techniques could be coded using the visual basic scripts (which Statsoft claim adds an additional 10, 000 data analysis and graphics functions to the industry standard syntax of Microsoft Visual Basic). However, it would have been nice if a bootstrap/jack-knife module was available, in addition to more smoothing options, and at least options to change the amount of smoothing within the LOWESS routine.

Graphic Facilities

The graphic facilities in STATISTICA are extensive and combine an extremely wide selection of 2 and 3-dimensional graphical displays (Figure 6). One of the more powerful aspects of STATISTICA, when compared to other statistical software, is the flexibility of graph customisation and the graphic management facilities. Every aspect of a STATISTICA graph can be edited (e.g. fitting additional functions, data smoothing, axis scaling etc) in place without having to return to the original graph dialog box. Once created, all STATISTICA graph documents contain all options, features, styles, information about the inserted, linked or embedded objects, as well as all the relevant data, and therefore can be edited or customised further (e.g., change fitting options, categorization settings, etc) without needing access to the original dataset. Graphs saved in files or pasted in a document in another application are ActiveX objects and therefore can be edited in place.

For example, a student could create all the graphs necessary for a particular analysis and then import them into a word processor. All editing can now be carried out in the word processor from simple formatting, such as deciding on colour styles and choosing appropriate axis scales, to adding categorisations (e.g. gender labels) without having to revisit the dataset. In addition, graphs can be set to update automatically whenever the dataset is amended.

A variety of brushing techniques for exploratory data analysis including data selection and identification methods, interactive labelling, marking, animated brushing, and value range and attribute-based brushing are available. The animated brushing routine would be particularly useful when motivating correlation and regression and in courses involving classification methods.

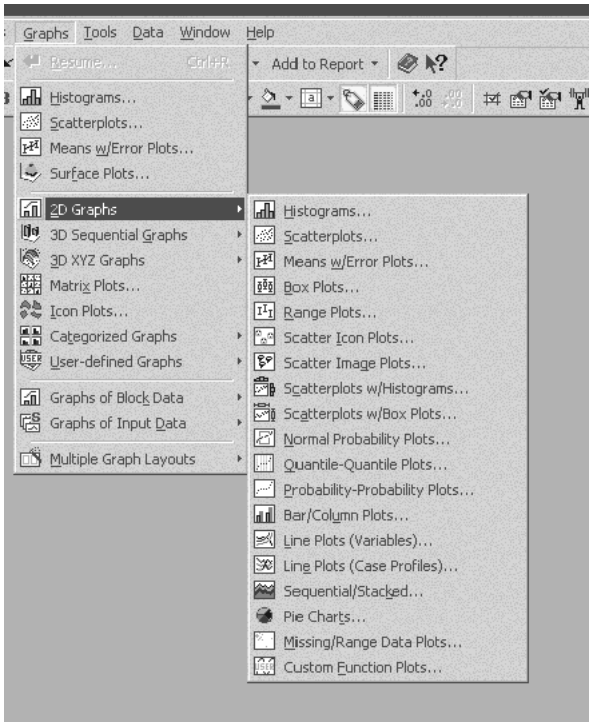

Figure 6

Output

STATISTICA generates output in three optional forms: workbooks, reports, and stand-alone windows.

Workbooks have become a standard in many statistical packages and are a useful way to manage output and graphics by using a hierarchy of tabs and folders.
Reports offer a more traditional way of handling output where each object is displayed sequentially in a word processor style document which can form the basis of a final report. All objects in the report (e.g. graphs, tables etc) are embedded and can be edited in place and a report can be saved in rich text or html format.
Stand-alone windows, the traditional method of displaying output in previous versions of STATISTICA, are still available. In this method, each piece of output is directed to a separate window, which is annoying as the screen becomes very cluttered making it very easy to lose important graphs and results.

For teaching purposes, using either the workbook or report format appears best. Students could collect the various pieces of output they deem necessary in a workbook, which could then be redirected to a report. Alternatively, a report could be generated on immediately and edited afterwards in a word processor.

Help files

There were no manuals provided with the software on review, however STATISTICA has an extensive electronic manual, which provides comprehensive reference information on all procedures, facilities, and options of the software. The electronic manual is well written and includes many examples. In addition, there is an excellent free electronic textbook available on the STATISTICA website in addition to a library of contributed visual basic routines.

Conclusion

STATISTICA is a very fast and powerful statistics package and is very much a windows based application. Statisticians who currently use command driven applications will probably dislike it while windows friendly statisticians will find it very easy to use and extremely customisable. The graphic capabilities are very impressive especially the ability to edit virtually every component of the graph and the fact the data is always available within the graph. It is not surprising that the software has won many awards from the business community given its ease of use and its power.

As a teaching tool, STATISTICA can easily provide the necessary statistical features needed for university undergraduate statistics courses and for most routine analyses undertaken by graduate students in disciplines such as the sciences, engineering, psychology and business. I am not convinced however of the usefulness of STATISTICA at present as a statistical research tool as this will depend on how many users decide to write (and share) code in the future. Given the amount of freely available code for S-Plus and R it is difficult to imagine STATISTICA being able to compete as a research tool.

In conclusion, STATISTICA 6 is a very easy to use and comprehensive analysis tool kit, very fast and powerful, a good teaching tool but not recommended yet as a dedicated research tool.

StatSoft comments on the *STATISTICA* 6 review from Helene Dubus

Further to the pricing information comments in the introduction
- The academic price for *STATISTICA Base* is £425.
- Optional add-ons are available at prices ranging from £210 to £315 (academic prices).
- *STATISTICA Neural Networks* is available as an add-on or as a stand-alone program.
- The Student Edition of *STATISTICA* 6 will soon be available for approx. £50.
- Please call StatSoft Ltd on 01234 341226 for Site License information.

Data Handling: As mentioned in the review, *STATISTICA* can read data from most standard spreadsheet formats. It also features a powerful visual Query Wizard that can be used to import data from any ODBC-compliant database.

Conclusion: Overall, this is a fair assessment of the suitability of *STATISTICA* for use as a teaching tool and the breadth of statistical and graphical tools illustrated in the review also make the package ideal for use in many areas of research. For historical reasons, *STATISTICA* has often not been be the first choice for mathematical and statistical researchers, but with the addition of Visual Basic, we feel it is now well-placed for use in this type of custom development work.