

STATGRAPHICS VERSION 5: OVERVIEW & TUTORIAL GUIDE

**Professor Robert Nau
Fuqua School of Business
Duke University
February 2005**

CONTENTS

Topic	page
1. GENERAL FEATURES	1
2. ENTERING DATA	5
3. RUNNING PROCEDURES AND SAVING WORK: IMPORTANT SECRETS	7
4. DATA TRANSFORMATIONS AND MATHEMATICAL OPERATIONS	11
5. TRANSFERRING REPORTS AND GRAPHS TO YOUR WORD PROCESSOR	13
6. TUTORIAL: THE MULTIPLE-VARIABLE ANALYSIS PROCEDURE	14
7. THE ONE-WAY ANOVA (Analysis Of Variance) PROCEDURE	16
8. THE MULTIPLE REGRESSION PROCEDURE	17
9. SNAPSTATS AND AUTOMATIC FORECASTING	20
10. ON-LINE DOCUMENTATION	22

STATGRAPHICS VERSION 5: OVERVIEW & TUTORIAL GUIDE

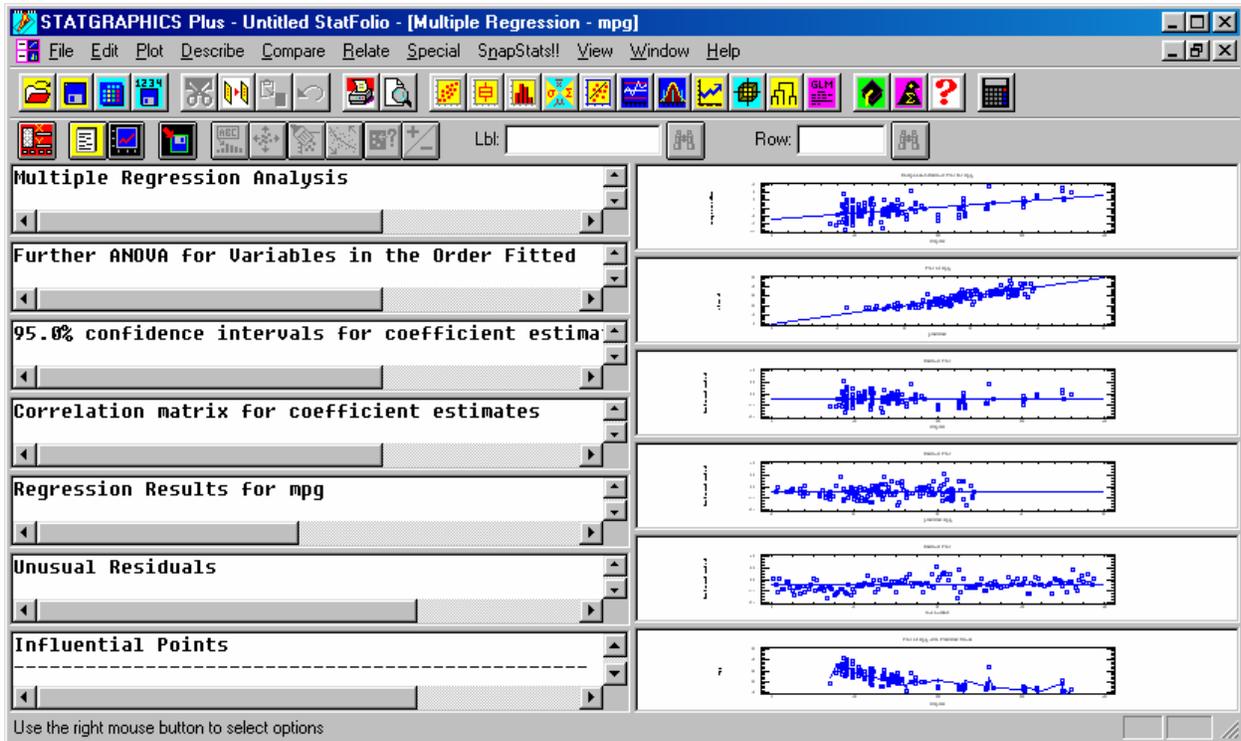
1. GENERAL FEATURES

Statgraphics Plus for Windows is Duke's official site-licensed PC statistics package for general use. Statgraphics is a full-featured statistics package that has much more modeling power and flexibility than spreadsheet add-ins such as Regress, while it is easier to use and more graphics-oriented than better-known large packages such as SAS or SPSS. Whereas many stat programs are grab-bags of analytical and graphical subroutines that fill your desktop with unorganized reports and graphs in overlapping windows, Statgraphics arranges your statistical activities into structured "analyses," "folios," and "galleries."

Analyses: An "Analysis" is a single window that is created whenever you run a statistical procedure, and it contains *all* the reports and graphs produced by that procedure. For example, the Analysis window for multiple regression might contain a collection of text reports showing coefficients, residuals, forecasts, etc., plus a corresponding collection of graphs. Each separate report or graph constitutes a "pane" within the analysis window. You can view all the panes simultaneously (which is often convenient when there are multiple graphs), or you can "maximize" any single pane to study it in more detail. *Double-clicking* on any pane causes it to be maximized, and double-clicking it *again* returns it to its original size and location. You may have many different analyses active on your desktop at any one time—e.g. some descriptive statistics procedures as well as a variety of different multiple regression or forecasting models that you might wish to compare. You can switch back and forth between different analyses by maximizing or minimizing them in the usual Windows fashion. (Just click on the icons in their upper left and right corner, etc.). You can also print the current pane or all panes—although it is recommended that you not use the "print all panes" feature indiscriminately. It is usually better to use the "Statgallery" feature (described below) to collect several graphs on a page before printing, and/or to export key reports and graphs to your word processor (as described below) for inclusion in a document summarizing your analysis.

All analyses are "live" at all times: if you edit the data in the data file, or load a new data file containing variables with the same names, all the analyses currently on the desktop can be updated to reflect the changes. (The Edit Preferences settings determine whether the updating is automatic for all analyses or only for the currently active icon.) This is a powerful feature which has implications for how you may wish to organize your data. For example, suppose you have data on sales and advertising in a number of different markets and that you wish to perform the same kind of analysis (regression or whatever) on each market. Then you should store the data for each market in a separate file, and use the same variable names (say SALES and ADVERTISING) within each file. Once you have run the analyses for one market, you can then repeat the same analyses for another market by simply loading a different data file.

Example of an “Analysis Window” for multiple regression analysis: the lower toolbar controls the Analysis options such as selecting reports and graphs to display. Double-click on one of the “panes” to maximize a text report or graph):



Statfolios: A “Statfolio” file contains a collection of all the analyses that have been performed on a particular data file in a particular session. When you save a “Statfolio,” you are saving a snapshot of your current Statgraphics session so that you can return to it later and find everything exactly as you left it. This is another powerful feature, because it means that you do not have to remember or reconstruct long sequences of operations in order to reproduce the results of a previous session or to obtain similar results for a different data set.

The StatAdvisor: Statgraphics includes a very useful “StatAdvisor” feature: at the bottom of every text report, a verbal summary of your results is given which includes specific comments on the meaning and significance of the statistics in the report. This verbal summary is generated by a built-in statistical expert system. It is not guaranteed to give you the last word on your analysis, but it does help to provide to some perspective on your results and alert you to obvious problems. There is also a “StatAdvisor” icon at the bottom of your Statgraphics desktop. If you double-click this icon, it opens a window in which the StatAdvisor summary is displayed for the text or graphics pane that is currently selected. (After opening the StatAdvisor window, you can select different text or graphics panes by clicking on them, and the text in the StatAdvisor window will change appropriately.) The StatAdvisor window can be used to obtain generic explanations of graphs, although it does not specifically comment on the details of each graph—you’ll have to rely on your own pattern-recognition capabilities for that!

Example of a “StatAdvisor” report for multiple regression analysis:

The output shows the results of fitting a multiple linear regression model to describe the relationship between mpg and 7 independent variables. The equation of the fitted model is

$$\text{mpg} = 27.8578 + 0.431885 \cdot \text{accel} + 0.436845 \cdot \text{cylinders} + 0.0287842 \cdot \text{displace} - 0.0554725 \cdot \text{horsepower} + 0.000622572 \cdot \text{price} - 0.0123731 \cdot \text{weight} + 0.280967 \cdot \text{year}$$

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 77.2534% of the variability in mpg. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 76.1321%. The standard error of the estimate shows the standard deviation of the residuals to be 3.59917. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 2.65875 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0.4799, belonging to cylinders. Since the P-value is greater or equal to 0.10, that term is not statistically significant at the 90% or higher confidence level. Consequently, you should consider removing cylinders from the model.

Parameter	Estimate
CONSTANT	27.8578
accel	0.431885
cylinders	0.436845
displace	0.0287842
horsepower	-0.0554725
price	0.000622572
weight	-0.0123731
year	0.280967

Source	Sum of Squares
Model	6247.31
Residual	1839.47
Total (Corr.)	8086.77

R-squared = 77.2534 percent
R-squared (adjusted for d.f.) = 76
Standard Error of Est. = 3.59917

Data files: Statgraphics stores data in its own proprietary format, but it can also read data files created by other programs. In particular, it can read text files with various delimiting characters, spreadsheet (Excel and Lotus 123) files, and ODBC database files (Access, Oracle, etc.). Data can also be copied or linked to the Statgraphics data spreadsheet from other spreadsheet programs via the usual Windows copy, paste, and paste-link operations.

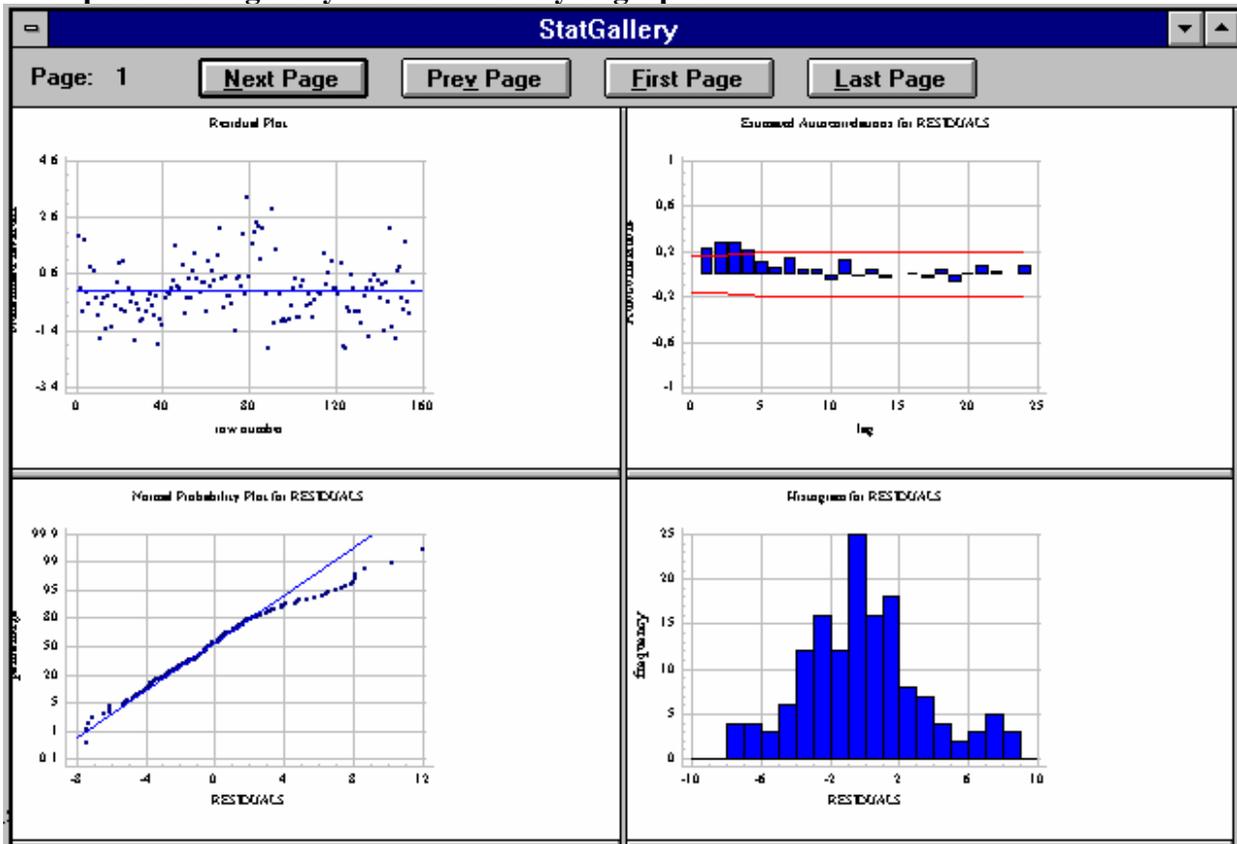
Menu structure and toolbars: Statgraphics has a menu bar at the top of the screen with standard pull-down menus like any other Windows application. The top-level menu is organized in a task-oriented rather than tool-oriented way: the choices are “Plot,” “Describe,” “Compare,” “Relate,” etc. These lead to pull-down menus with more specific choices such as descriptive statistics, regression, ANOVA, etc. Immediately below the main menu bar is a toolbar containing icons for commonly used menu operations (e.g., opening and saving data files, copying and pasting, and running commonly used procedures like summary statistics and regression). This toolbar is merely for convenience: all of its icons correspond to procedures you can select from the menu, so you don’t have to memorize what they do. When you run a particular statistical analysis procedure from the menu, an “Analysis Window” will appear (as described above) with its own toolbar. The Analysis Window toolbar contains icons for essential operations like choosing which reports and graphs to display, annotating and rotating graphs, and identifying point values.

Secrets: Statgraphics works similarly to other Windows applications and is reasonably self-explanatory in its statistical functions, as long as you have at least a vague idea of the kind of analysis you want to perform. However, there are **three non-obvious secrets** you need to know in order to use the program effectively. These are (1) how the icons on the **Analysis Window toolbar** (mentioned in the preceding paragraph) are used to obtain the full range of text and graphics reports for each procedure (unless you use these toolbar icons, you will not necessarily see all the available output when you run a procedure); (2) how the **right mouse button** is used to obtain access to procedure and graphics options (unless you use the right mouse button, you will never discover many of the program's most interesting features); and (3) that Statgraphics uses **three different kinds of files** in which to store your work, namely a DATA file that contains your raw data, a STATFOLIO file that contains your "live" analyses, reports and graphs, and a STATGALLERY file that may contain additional saved graphs from previous analyses. These secrets are described in more detail in the "Running Procedures" section below.

Program modules: Like many stat programs, Statgraphics is modular in design: there is a "base" module containing the most commonly used procedures (including regression and ANOVA), a new quick-analysis option called Snapstats!, and a series of additional modules for Quality Control, Experimental Design, Time Series Analysis, Multivariate Methods, and Advanced Regression. The additional modules appear on the "Special" menu in Statgraphics. The Time Series Analysis module is particularly innovative and includes additional regression options (e.g., the ability to analyze hold-out data and/or specify ARIMA error structures) that are not available in the standard multiple regression procedure.

The Statgallery: Another very nifty feature for organizing text and graphics output is the "Statgallery." The Statgallery is a gallery of pages on which you can paste an array of reports and graphs in any rectangular configuration up to 3x3 (i.e., 3 graphs or reports across the page and 3 down). This feature enables you to collect reports and graphs from one or many procedures and organize them onto a single page or an album of several pages. Whenever a report or graph is on the screen, the "Copy to Statgallery" option is available by hitting the right mouse button. After selecting this option, maximize the Statgallery (by double-clicking its icon or clicking its maximize handle) and click with the *right* mouse button on the pane where you want the graph or report to be pasted. Finally, choose "Paste" from the pop-up menu that appears. (Alternatively, you can choose "Paste Link" for a live link back to the original analysis.) You can view the current contents of the Statgallery at any time during your session, and it can be saved to a file. When the Statgallery is being viewed, its configuration (i.e., number of rows and columns) also can be changed as a right-mouse-button option. If you get in the habit of using the Statgallery, you will greatly reduce the number of pages of paper you print out and you will also be able to better organize your output for viewing by others. Small graphs are usually as effective as big ones, and the more graphs you can fit on one page, the better.

Example of a Statgallery with a 2x2 array of graphs:



2. ENTERING DATA

When you first enter the program, a “Statwizard” pops up to offer you help in getting started with data entry and analysis. You can hit Cancel and “Yes” to exit if you don’t want any help, at which point you will see an empty data spreadsheet with the name “untitled.” (If you don’t see the spreadsheet yet, look for a colorful spreadsheet-like icon called <untitled> at the bottom of the screen and click its maximize handle.) Before you can do any analysis, you must enter or load some data onto the spreadsheet. You can do this in one of the following ways:

- a. Enter some data manually into the Statgraphics spreadsheet:** just enter data as you would in any spreadsheet program. Each column on the spreadsheet corresponds to a variable in the data file, and the default column/variable names are Col_1, Col_2, etc. To assign a more descriptive name to a column, or change its other attributes, highlight the entire column by clicking on the *name bar* at the top (i.e., by clicking on the Col_1, Col_2, or whatever). Then click the *right mouse button* to bring up a dialog box in which to change column attributes, and select the Modify Column option to get a dialog box for changing the column name or formatting. When you are finished entering data, you should choose the “File/Save Data File” option from the menu bar to save the data before proceeding. At this time you can give the file a name that will henceforth appear alongside the spreadsheet icon.

b. Copy some data from another spreadsheet program onto the Statgraphics spreadsheet: First go to your other spreadsheet program (e.g., Excel), highlight the range of data you wish to copy, and select “Edit/Copy” from the menu bar (or hit Ctrl-C) to copy it to the clipboard. Then return to Statgraphics, maximize the spreadsheet (if necessary), position the cursor in the upper left cell (or wherever you wish the first column of data to be copied) and select “Edit/Paste” from the menu bar (or hit Ctrl-V). The data will be copied into the Statgraphics spreadsheet exactly as it appeared before. Alternatively, you can select “Edit/Paste Link” to create a dynamic link from the Statgraphics spreadsheet to the original spreadsheet file. That latter method should be used if the data is periodically going to be edited or updated in the spreadsheet file. (Data which has been linked rather than copied will appear in a grey rather than black font.) Alas, Statgraphics does not look for variable names when data is copied from the clipboard: you still have to enter your own variable names for the columns as described in the preceding paragraph.

c. Open an existing data file: Select the “File/Open Data File” command from the menu, then locate and select the file you wish to open. Statgraphics data files are identified by the extension .SF or .SF3 following their filenames. You can also open data files in ASCII text form (either tab-delimited, comma-delimited, space-delimited, or fixed format) or in Excel or Lotus format, with an option to read variable names from the first row in the file. (To see non-Statgraphics files listed, you must change the default file to to “Excel files” or “All files”.) **If you wish to import data from Excel into Statgraphics, including variable names, the best way is to arrange the data columnwise on the worksheet with variable names in the first row, save the worksheet as an Excel worksheet (.XLS) file and close it, then go to Statgraphics and use the “File/Open Data File” command to open it.** Variables names may be up to 32 characters in length, but it is recommended that you use short names (say, 10 characters or less) to avoid truncated names on formatted reports and to simplify typing of names when using data transformations. Note: if you don’t see your Excel file listed in the directory where you saved it, be sure you have changed the file type in the Open File dialog box to “Excel files.” If you get an error message saying “unable to open file,” be sure that the Excel file is not still open in your Excel window. **You must close the file in Excel before opening it in Statgraphics.**

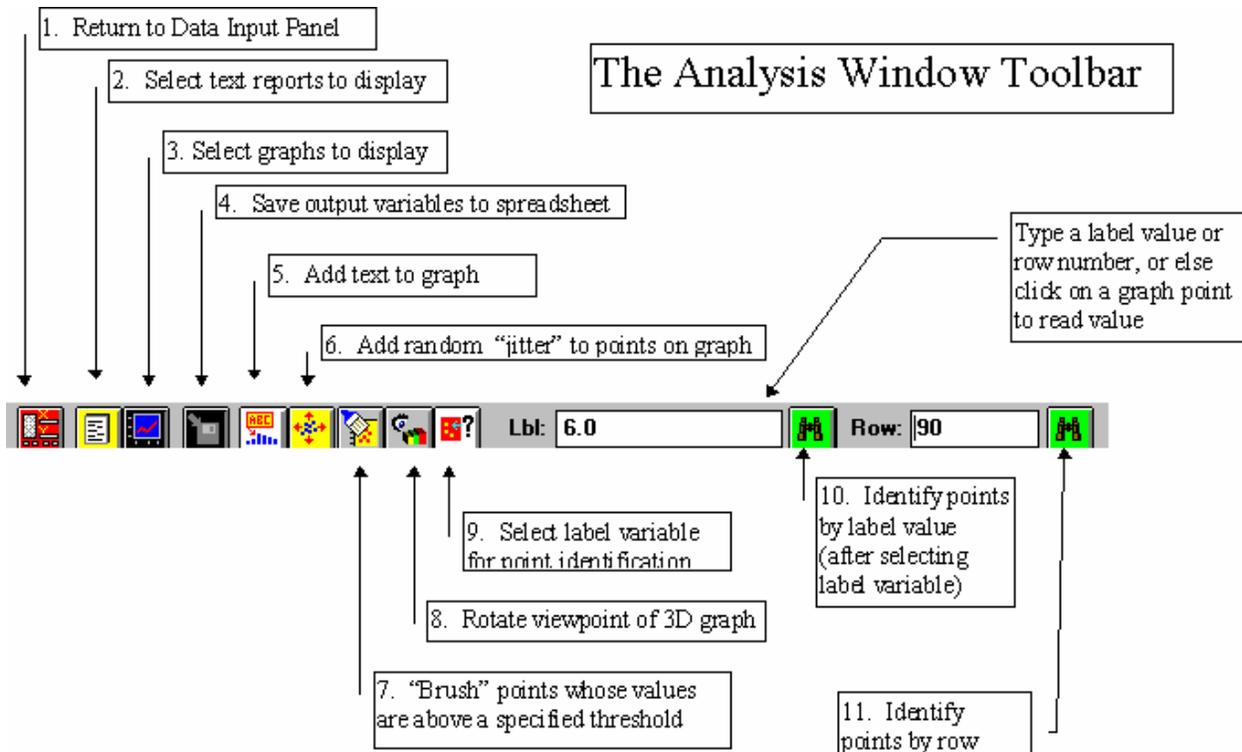
d. Open an existing statfolio: As noted above, a “Statfolio” file contains the “live” analyses that have already been specified for a particular data file. Statfolio files are identified by the extension .SGP following their filenames. To open a Statfolio file, just select the “File/Open Statfolio” command from the menu bar, then locate and select the SGP file you wish to open. When you open the Statfolio file, it will usually open the corresponding data file that was last used, but you can also load a different data file later.

3. RUNNING PROCEDURES AND SAVING WORK: IMPORTANT SECRETS

Data input: After you have entered or loaded data on the spreadsheet, you can select any graphical or statistical analysis procedure from the menu system. The procedures all have the same general layout. First you are presented with a **Data Input panel** on which to specify the variables to be analyzed and the mathematical transformations and/or selection criteria (if any) to be applied to them. There may be one or more input fields in which variables are to be specified. At the left of the screen you will be shown a list of the variables currently on the spreadsheet to choose from: to enter one of these variable names into a data input field, highlight its name and then click the arrow icon that appears next to the desired field. Alternatively, you can just type the name of the variable, or some transformation therefore, in the field. For example, if you want to use X-squared rather than X in the analysis, you can enter X^2 as the variable name in the input field. If you want to use X plus Y, just enter $X+Y$. If you want to use X lagged by one period, enter $LAG(X,1)$, and so on. (A complete list of available mathematical functions is shown if you click the **Transform** button at the bottom of the Data Input panel. The Transform dialog box also includes an option to **Display** the first few rows of a variable or transformation thereof—this is often useful for verifying that the data looks like what you think it does. See the section on “Data Transformations and Mathematical Expressions” below for more details on mathematical operations.)

➡ **Secret #1: The Analysis Window toolbar:** After you have filled in the Data Input panel for a procedure, click OK and the analysis will be performed. The Analysis Window for that procedure will then appear on the screen, with its own toolbar at the top. Initially, the Analysis Window will contain a few default text reports and graphs—but these are usually not all the reports and graphs available. To see what additional *text reports* are available, click the *second* icon on the Analysis Window toolbar (a white text page on a yellow background) and you will see a list of all reports produced by the procedure, with check boxes to select them. To see what additional *graphs* are available, click the *third* icon on the toolbar (a blue line graph on a black background), and you will get a similar list of graphs with check-boxes. I suggest that you normally click both of these icons and choose the “All” option in each case to see all available reports and graphs. After selecting more reports and graphs to display, the Analysis window will now be filled with many different “panes,” with text panes on the left and graphics panes on the right. Double-click on a pane to maximize it (i.e., blow it up to fill the screen), then double-click again to shrink it back to its original size. To print some or all of the panes, click the printer icon on the Statgraphics main tool bar at the top of the screen.

The Analysis Window Toolbar and a description of its icons is given on the following pages:



Here is a complete list of Analysis Window toolbar icons and their functions:

1st icon: return to the Data Input Panel to change the variable specifications and/or selection criteria. This is useful if you want to change any variable specifications in the middle of the analysis (e.g., add or delete a variable from a regression analysis or a multiple-variable plot).

2nd icon: select *text reports* to display—*always click this to see what's available!*

3rd icon: select *graphs* to display: *ditto!*

4th icon: save output variables from the procedure as additional columns on the spreadsheet. *Important: this feature enables you to pass modeling results back and forth between procedures, if this turns out to be necessary for more advanced analysis or plotting. For instance, if you want to draw an autocorrelation plot or normal probability plot of residuals from an ordinary multiple regression analysis, which doesn't include these plots among its defaults, you can use this tool to save the residuals to the spreadsheet, then go to the appropriate plotting procedure to plot them. (This can all be done in a matter of seconds with just a few mouse clicks.)*

The remaining icons are *only* activated when a *graphics* pane is *maximized* on the screen.

5th icon: add text to graph

6th icon: add random “jitter” to points on graph—this is useful if some variables are discrete-valued resulting in data points plotted on top of each other. Jittering spreads them out.

7th icon: “brush” points on graph (change colors of points for which some variable lies between lower and upper bounds determined by two “sliders”)

8th icon: rotate viewpoint (3D graphs only)

9th icon: select label variable for point identification. This icon works in conjunction with the “binoculars” icon immediately to its right (by the “Lbl” box). When you click this icon you are asked to select a variable containing a data label—normally a variable that takes on only discrete values. After doing this, click on any point in the graph: you will see its label value appear in the “Lbl” box. If you now click the binoculars icon next to this box, *all* points having the *same label value* will be highlighted on the screen. (Alternatively you can just type a label value in the box and then hit the binoculars icon.)

The second binoculars icon, next to the “row” box, works similarly, except that it highlights points with a specified *row number*. Click on a graph point to see its row number displayed, then click the binoculars icon to highlight all other points associated with the same row.

➡ **Secret #2: right mouse button = procedure and graphics options:** Many essential operations in Statgraphics—namely the *options* associated with statistical and graphical procedures—are controlled exclusively by the *right* mouse button. Whenever a text report or graph is on the screen, clicking the right mouse button will usually bring up a menu of options you can change. For text reports (particularly the “Analysis Summary” report for a statistical procedure), the options are those of the statistical procedure itself—e.g., what type of regression or forecasting to perform, what tolerances or confidence limits to use during estimation, etc. Therefore, **whenever you look at the “Analysis Summary” report for a procedure, you should immediately hit the right mouse button to see what other analysis options (if any) are available for that procedure.** You will then see a menu of available options for the procedure. For example, if you have run a multiple regression, one of the available options is to re-run it as a *stepwise regression*. Thus, stepwise regression is an available procedure even though you didn’t see it on the main menu. Many other variations on standard procedures—e.g., different kinds of ANOVA or forecasting models—are also controlled by the right mouse button in this way. Other text reports (besides the “Analysis Summary”) may also have their own right-mouse-button options.

For *graphs*, the right-mouse-button options typically control the way in which the output of the procedure is plotted. There are usually many different kinds of options available behind any particular graph. Before hitting the right mouse button, click on the feature of the graph that you would like to change—e.g., the title, the axis labels, the axis scale, the points or lines associated with a particular variable—then click the right mouse button and choose “Graphics Options.” This brings up a dialog box with tab pages for changing different elements of the graph—the tab page that corresponds to the selected graph feature usually will appear first. If you are running

Statgraphics on your own PC, you can usually save any changes that you make to the options for a particular graph type so that they will apply to all similar graphs in the future. To save the settings of the current graph after you have modified them, select the “Profile” tab from the “Graphics Options” box, then specify “User #1” (or whatever) and “Make Default” in the dialog box which appears, and use the “Save As” option to save the profile. (You can also save additional sets of graphics settings and reload them later under headings User #2, User#3, etc.) Modifications that you make to graphics settings during a session will be applied to any new analyses that are run from that point onward, but they do not apply retroactively to other graphs in analyses that have already been performed.

➡ **Secret #3: Four kinds of files:** After you have run some procedures in Statgraphics and you wish to save your work, you need to be aware that there (potentially) *four different files* to keep track of. First, there is the **DATA** file in which your raw data is stored. Sometimes this file is modified by the results of your analysis—e.g., if you have entered new data or created new variables on the data spreadsheet via mathematical transformations, or if you have saved some of your procedure outputs such as forecasts and residuals on the data spreadsheet. The data file has an extension SF or SF3 in its DOS file name. Second, there is your **STATFOLIO** file. This file contains the details of all the “live” statistical analyses on your Statgraphics desktop—i.e., it contains a snapshot of your session. It has the file extension SGP. Third, there is the optional **STATGALLERY** file which contains any graphs that you have previously copied into the Statgallery from the statistical procedures. (It is not necessary to use the Statgallery, but it sometimes provides a convenient way to organize and display graphs produced by different procedures, as described above.) The Statgallery file has the extension SGG. Fourth, there is the optional **STATREPORTER** file, which is a rich-text-format file (with extension RTF) into which you can copy entire analyses (all reports and graphs) with a single mouse-click, so as to leave a complete audit trail which can be imported directly into Microsoft Word. To use the Statreporter feature, just click the right mouse button while an analysis is open and then click the “Copy Analysis to Statreporter” option on the menu. This will add all the reports and graphs from the active analysis window to whatever is already in the Statreporter window. You should then go to the “File/Save As” menu to save the current contents of the Statreporter to an rtf file. (If you are going to use the Statreporter, it is a good idea to save its contents frequently so that you will have an audit trail in the event of a program crash. You can always replace the existing contents rather than creating a new file.) Later, you can open the Statreporter file with Microsoft Word and edit its contents (e.g., copying and pasting individual reports and graphs to other documents), or you can just add the whole thing as an appendix to your writeup of your analysis.

Most often, when you are working on a statistical project, you will have a *single data file* and a *single Statfolio file* and (optionally) you may also have a *single Statgallery file* and/or a *single Statreporter file*. If you hit the “Save” button on the main toolbar, this will save your current Statfolio file and it will also prompt you to save the data, Statgallery, and Statreporter files if they have been modified. Often you will wish to give the *same filename* to all the files, although they will have different file types and hence different extensions. Thus, for example, if you are working on a project called SALESFORECAST, you might end up with three files called SALESFORECAST.SF3 (the data file), SALESFORECAST.SGP (the Statfolio file), SALESFORECAST.SGG (the Statgallery file), and/or SALESFORECAST.RTF (the Statreporter file).

VERY IMPORTANT: If you want to move your work to a different computer or network directory or share it with someone else, you need to move **BOTH** the data and Statfolio files, as well as the Statgallery and/or the Statreporter files if you have used either of those two optional features.

If you have moved the files to a different location, then when you first open the Statfolio file, you may need to respecify the relative location of the corresponding data and Statgallery files. It is also possible to have *different* data files or Statgallery files associated with the same Statfolio file, and vice versa—this is one of the flexible and powerful features of Statgraphics. For example, you could run different sets of procedures on the same data file, stored in different Statfolios. Or, you could have different data files with the same generic variable names (say, sales data from different regions or time periods) that are analyzed with a common set of procedures in a single Statfolio file. Statgallery files are somewhat independent—you could have a Statgallery file that contains graphs produced from various data sets and/or Statfolios.

How to avoid losing your work: As with any other application, you should save your work frequently. Fortunately the Statfolio feature makes it easy to save and restore your Statgraphics session—just hit the “Save” button on the main toolbar to save the current Statfolio. (It is also a good idea change the name periodically using “Save As” to keep several generations of the file.) Try not to let analyses multiply like rabbits on your Statgraphics desktop: each new analysis requires its own icon and monitoring that consume scarce “system resources.” While it is fine to have a dozen or so analyses in the same Statfolio, users who encountered problems in the past were often found to have *several dozen* analyses (e.g., 25 different regression models) active at the same time, so use at least a little restraint. There is (or was) also a limit of 95 *total* graphics panes (on all pages) in the Statgallery at any one time—the system could hang if you try to cut-and-paste beyond this limit. If you want to run extremely large numbers (i.e., dozens) of different analyses, it will be easier to keep track of them anyway if they are grouped in several different Statfolio files. Remember that it is easy to modify existing analyses (e.g., by changing the variable specifications or model options) and then re-save the Statfolio under another name.

4. DATA TRANSFORMATIONS AND MATHEMATICAL OPERATIONS

If you want to use a transformation of an existing variable or variables as the input to a statistical procedure, you do not need to create a new variable for this purpose: wherever you are prompted to enter a variable name as input to a procedure, you can simply enter a mathematical formula to perform the transformation on-the-fly. For example, if you want to use **SALES** divided by **CPI** as the input to a procedure (assuming these are two existing variables), you can just type **SALES/CPI** as the variable name. A "Transform" button usually appears at the bottom of input screens where variable names are entered: this provides further assistance in selecting mathematical functions and testing the formula that you are creating.

Creating a new variable: If you do wish to create a new permanent variable on the data spreadsheet, computed from a mathematical formula or a transformation of existing variables, proceed as follows:

- Maximize the data spreadsheet icon on the Statgraphics desktop.
- Click on the title bar of a new (empty) column, which highlights the entire column.
- To fill the column with data generated by a “live” formula, click the *right* mouse button and select **Modify Column** from the menu. Enter a name for the new variable in the name field at the top, then click the **Formula** radio button at the bottom and type the desired mathematical expression in the dialog box.
- Alternatively, if you wish to *hard-code* the new data instead of using a live formula, use the **Generate Data** right-mouse-button option instead of Modify Column, and enter the desired mathematical expression in its dialog box. Later, use the Modify Column option separately to assign a name to the column.

Examples of valid mathematical expressions (X and Y denote arbitrary variable names):

- Usual arithmetic operations: $X+Y$, $X-Y$, $X*Y$, X/Y , X^Y
- Usual nonlinear functions: **LOG(X)**, **EXP(X)**, **LOG10(X)**, **EXP10(X)**, **SQRT(X)**, **ABS(X)**
- Statistical functions: **AVG(X)**, **SD(X)**, **VARIANCE(X)**
- Count from a to b in steps of size c : **COUNT(a,b,c)**
- Index variable (1,2,3,...) of the same length as X: **COUNT(1,SIZE(X),1)**
- Dummy variable (e.g. "select" condition) for occurrences of X greater than 2: $X>2$
- Similarly: $X=2$, $X>=2$, $X<>2$, etc.
- Logical "and" (ampersand): $X>2\&X<10$
- Logical "or" (vertical bar): $X<2|X>10$
- Logical negation (tilde symbol—negated expression must be in parentheses): $\sim(X<2)$
- Lag the variable X by 2 periods: **LAG(X,2)**
- First difference (period-to-period change) of X: **DIFF(X)**
- Seasonal difference of X with seasonal period = 12: **SDIFF(X,12)**
- Generate 100 uniform random numbers between 0 and 1: **RUNIFORM(100,0,1)**
- Generate 100 normal random numbers with mean 0 and standard deviation 1: **RNORMAL(100,0,1)**
- Similarly: **REXPONENTIAL**, **RGAMMA**, **RLOGNORMAL**, **RINTEGER**, **RWEIBULL**
- Cumulative normal probability of observation x : **NORMAL(x, mean, stddev)**
- Value of x which has cumulative normal probability equal to p : **INVNORMAL(p, mean, stddev)**
- Similarly: **CHISQUARE(x,df)**, **INVCHISQUARE(p,df)**, **STUDENT(x,df)**, **INVSTUDENT(p,df)**, etc.

- Subset of observations of X which meet a logical condition (otherwise treated as missing):
SELECT(X, *logical condition*)

You can also build compound expressions (e.g., $(X+2*Y)^{(1+Z)}$) using parentheses where necessary to force the desired order of calculation. For a complete list of built-in functions, hit the "Transform" button that appears at the bottom of dialog boxes where variable names and expressions may be entered. Within the "Transform" dialog box, you can hit the "Display" button to see what the first few calculated values look like, just to be sure your syntax is correct.

5. TRANSFERRING REPORTS AND GRAPHS TO YOUR WORD PROCESSOR

Pasting Statgraphics text reports into word processing documents: You can paste text reports from Statgraphic into your word processing document by copying them to the clipboard (i.e., using the Edit/Copy command or hitting Ctrl-C while the text report is maximized on the screen), and then pasting them from the clipboard into the document in the usual way (using Edit/Paste or Ctrl-V). This method transfers the reports as “Windows objects” that cannot subsequently be edited in the word processor. Another option is to first copy-and-paste one or more text reports into the "comments" notepad that is part of your Statfolio in Statgraphics. (Look for the notepad icon at the bottom of your Statgraphics desktop.) Later, you can copy-and-paste part or all of the contents of the comments notepad to your word processing document. Under the latter approach, the reports are copied as plain text that can subsequently be edited or formatted in your word processor. If you are planning to copy a lot of results into your report, the best approach is to use the Statreporter feature described above—i.e., use the right-mouse-button option to copy entire analyses into the Statreporter, which is then saved as a separate rich-text-format (RTF) file. The latter file can be opened and edited directly with Word.

Pasting Statgraphics graphs into word processing documents: If you wish to copy and paste graphs produced by Statgraphics into a word processing document, you can this in various ways. One way is to simply use the Statreporter feature to transfer graphs as well as text reports in an RTF file that can be opened and edited in Word. If you just want to transfer one or two graphs, without going through the Statreporter, you can either use the Windows clipboard (Edit/Copy or Ctrl-C, etc.) or else you can use the right-mouse button Save Graph command, while the cursor is positioned over the desired graph. The clipboard approach transfers the graph as a Windows object which can be pasted into a document via the usual Paste command (or Ctrl-V). The Save Graph approach saves it as a Windows Metafile (WMF), which can then be inserted into a Word document via the Insert/Picture/From-file command. (At this point you must go find the saved file, remembering the name you gave it and the directory in which you placed it.) If the inserted graph appears too small (e.g., postage stamp size), click on it to select it and then use the Format/Picture/Size command on the Word menu to resize it. When copying graphs by either method, be sure that your graphics options specify a WHITE BACKGROUND for the graph, otherwise it will be very hard to print out in Word! When you print a graph while in Statgraphics, the background color usually doesn't print, so it doesn't matter whether it is black or white. But once a graph is copied to Word, what you see is what you get. Also, be sure that relative height and width are adjusted appropriately in Statgraphics before copy or saving the graph: what you see is what you will get. After inserting a graph into Word, you can make it bigger or smaller, but you can't change the “aspect ratio”—i.e., the proportions. *You may wish*

to try this out after drawing one of the graphs in the tutorial steps below—i.e., double-click the graph to maximize it, use the copy or save-graph command to copy it to the clipboard or save it to a Windows metafile, then paste or insert it into a Word document

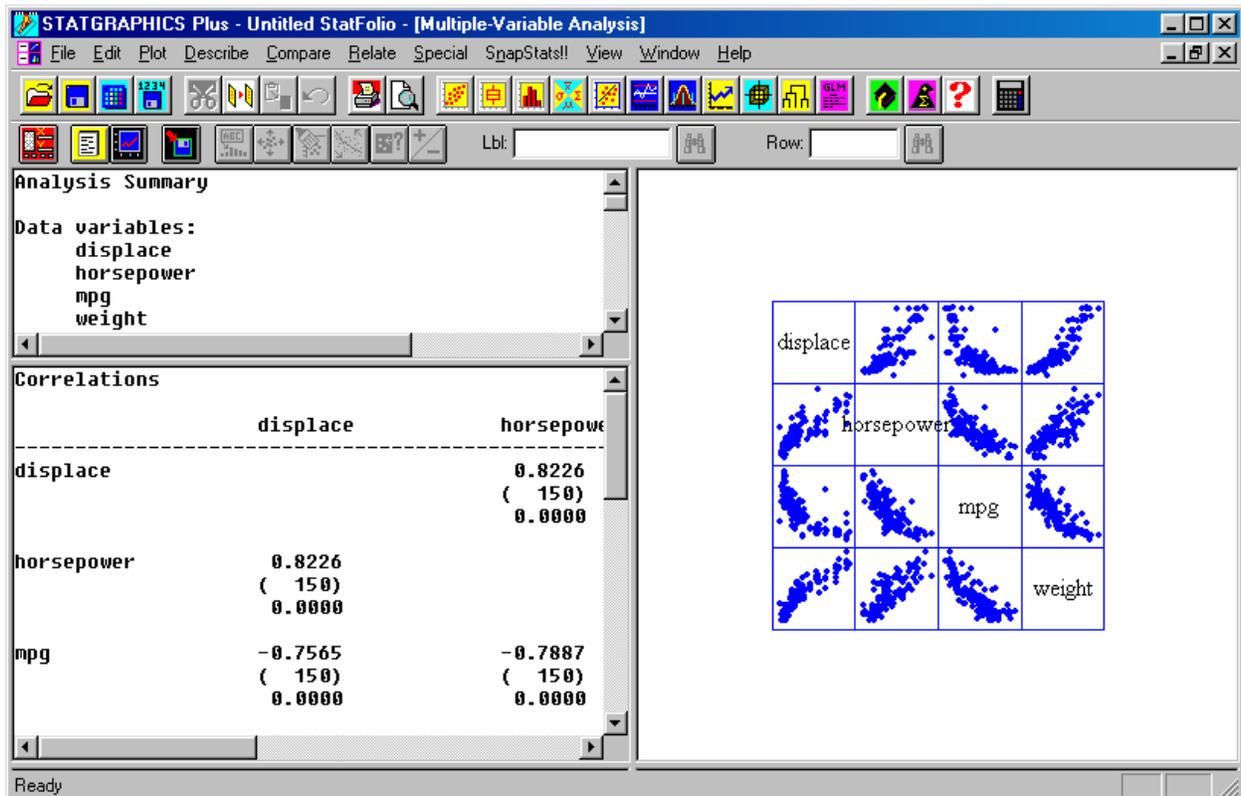
6. TUTORIAL: THE MULTIPLE-VARIABLE ANALYSIS PROCEDURE

To illustrate some of the basic operations described above, load the CARDATA file that is one of the sample data files included with Statgraphics. If you performed a complete installation on your PC, this file should appear in the directory **C:/Program Files/Statgraphics/STATGRAPHICS Plus 5.0/Data** that was automatically created at that time. Choose “File/Open Data File” from the main menu and open the CARDATA.SF file from that directory.

After the data file has loaded, select “Describe/Numeric Data/Multiple-Variable Analysis” from the menu. On the data input screen, you should see a list of all available variables in the file on the left-hand-side of the screen. Select the following four variables for analysis: *displace*, *horsepower*, *mpg*, and *weight*, then click “OK “. (To select a variable, just double-click on its name. Alternatively, you can click once on its name to highlight it, then click the arrow icon labelled “Data”.)

You should now see the screen fill with two text panes on the left—the Analysis Summary and Correlation Matrix reports—and a Scatterplot Matrix graphics pane on the right (i.e., a matrix of all 2-way scatter plots between pairs of variables, the visual counterpart of a correlation matrix).

Scatterplot matrix in Multi-Variable analysis (note that there are strong relationships between all pairs of variables, but the relationships are not necessarily linear):



As noted above, you can click the toolbar icons to display the other reports and graphs that are available. First click 2nd icon on the toolbar (white page on a yellow background) and click the “All” box on the list of *text reports*, then click “OK.” You should see a number of different “panes” appear in the window, containing additional text reports. You can “maximize” one of the panes for closer viewing by double-clicking on it. Try this now: double-click on the *second* text pane, which is the “Summary Statistics” report. This report will initially show seven different summary statistics for each of the four variables: count, average, variance, standard deviation, minimum, maximum, and sum. These are not the only summary statistics available: remember that many procedures have additional *options* which can be called up by hitting the *right mouse button*. Try this now: hit the right mouse button while viewing the Summary Statistics report: you should see a dialog box appear with a list of all the statistics that can be displayed (including median, mode, geometric mean, interquartile range, etc.). You can select as many or as few of these as you wish, then click “OK” to redraw the report. You can try the same thing with some of the other reports—i.e., maximize a report by double-clicking it, then hit the right mouse button to see what options are available. (Some may have no interesting options.)

By clicking the 3rd icon on the toolbar (the blue line graph on the black background) you could look at some of the other *graphs* that are available in this procedure. The other graphs, “star” and “sunray” plots and the “key glyph” that explains their coding, are not useful for this particular data set, however, so don’t bother.

Next, let’s try some of the graphics options. Maximize the scatterplot matrix by double-clicking on it. You should see some of the additional icons on the toolbar light up. First, you may want to make the points smaller or rounder or darker: *right-click* on the graph, choose Graphics Options, then choose the Points tab and use the controls to change the size and shape of the points and/or check the “fill point” box to fill them in. Next let’s try using the **brushing** feature: click the brushing icon (the 7th from the left on the toolbar). A bar with two “sliders” will appear at the top of the graph which you can use to vary a price threshold. All points corresponding to cars whose prices are *in between the values set by the sliders* will be colored red, while the others will be blue or black. If you leave one slider at the far left or right, the other slider determines a threshold above or below which the point color changes. By moving the second slider back and forth you can see how the other variables change with the price threshold. All of these variables are strongly related: the more expensive cars also tend to have larger (engine) displacements, more horsepower, more weight, and *lower* mpg.

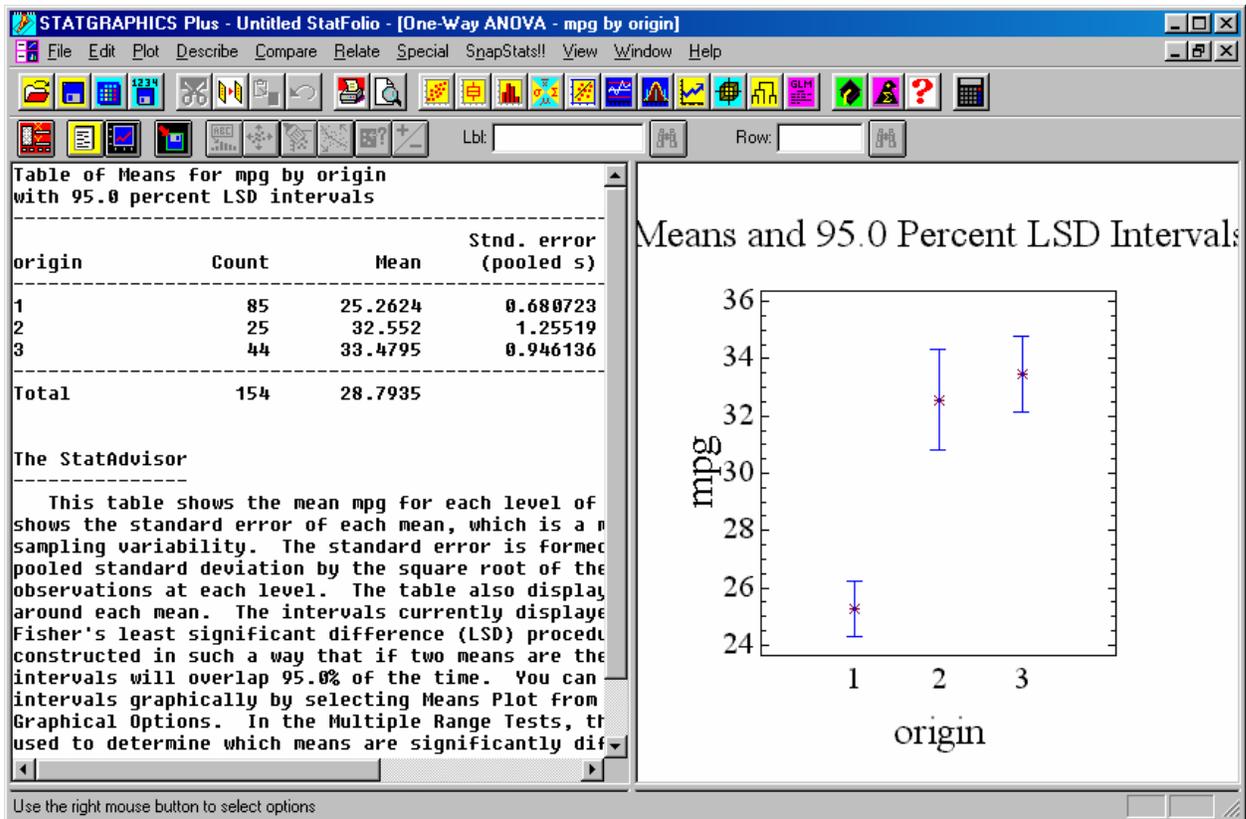
Now try the **point identification** feature. Click on the point identification icon (the 9th from the left on the toolbar). You will first be prompted to specify a variable containing the labels you want to identify. Choose “cylinders” as the label to use. Now move the cursor over the graph and click on some data point that looks interesting. You will see its label value (i.e., number of cylinders) and its row number displayed in the boxes at the top of the graph. If you now click on one of the **binoculars** icons, all the other data points with the same label value or row number will be highlighted in red. For example, if the data point you clicked on was for a 4-cylinder car, then clicking the binoculars by the “Lbl” box will highlight all the other 4-cylinder cars. You can also type a label value in the box (e.g. type 8 in the box and then click the binoculars to highlight all the 8-cylinder cars).

7. THE ONE-WAY ANOVA (Analysis Of Variance) PROCEDURE

The one-way ANOVA procedure is used for studying how the mean value of one variable varies across different levels or categories of some other, discrete-valued variable. (Note: in order to be used in the ANOVA procedure, the data for the dependent variable must be stored in a single variable, rather than separate variables for each category. A second variable is then used to store the “level codes” which distinguish the categories.) To run the One-Way ANOVA procedure, select “Compare/Analysis of Variance/One Way ANOVA” from the main menu. For starters, specify “mpg” as the dependent variable and “origin” as the factor (i.e., independent) variable. “Origin” is a categorical variable whose value is 1 for American cars, 2 for European cars, and 3 for Japanese cars in this data set. A one-way ANOVA of mpg on origin will determine whether there are significant differences in average miles-per-gallon between the American, European, and Japanese cars. After clicking OK on the Data Input panel, you will see couple of default reports and graphs, but these are not the only ones available. To see all the output that is available, click the 2nd and 3rd icons on the Analysis Window toolbar and select “All” text reports and “All” graphs. (You can check all the check-boxes or hit the “All” button at the bottom.) The first three graphs on the right-hand-side of the screen give different views of the data and model-fitting results. The first graph is simply a scatterplot by level code—i.e., the raw data. The second graph shows the means and 95 confidence limits, and the third graph is a box-and-whisker plot showing the medians and interquartile means (this is sometimes a more robust indicator of the location and dispersion of the data, particularly if the distribution is non-normal). The other graphs are various views of the residuals (deviations from the respective means). The text reports show some of the numbers behind the graphs (e.g., means, standard deviations, confidence intervals) as well as various statistical tests of differences between means. (Obviously there are some significant differences here.) After looking at some of the reports and graphs, minimize them so that you are once again looking at all the text and graphics panes simultaneously.

Now click the 2nd and 2rd icons again, and select only the “Table of means” report and the Means Plot” graph (i.e., un-check all the others). Your screen should now look like the picture on the following page. Now, just for the heck of it, change one or more of the variables in the model by clicking on the 1st icon on the toolbar to return to the Data Input panel. For example, change the independent variable from “mpg” to “year,” to investigate whether there have been significant changes in average mpg from year to year. Notice that the entire analysis is immediately redone with the new variables: the reports and graphs are updated. (What trends or changes do you observe in mpg over time?) Now try some more changes: for example, change the dependent variable to “horsepower” and the independent variable to “cylinders.” Except for the one oddball car with three cylinders, there appears to be clear relationship between cylinders and horsepower. In fact, it almost looks as though horsepower might be roughly proportional to number of cylinders—i.e., horsepower per cylinder might be roughly constant. You can test this hypothesis by changing the dependent variable to “horsepower/cylinders”. (That’s right: type “horsepower,” followed by a “slash” character signifying division, followed by “cylinders.”) This illustrates how you can use mathematical transformations of variables on-the-fly, without having to create new variables to store the transformed values.

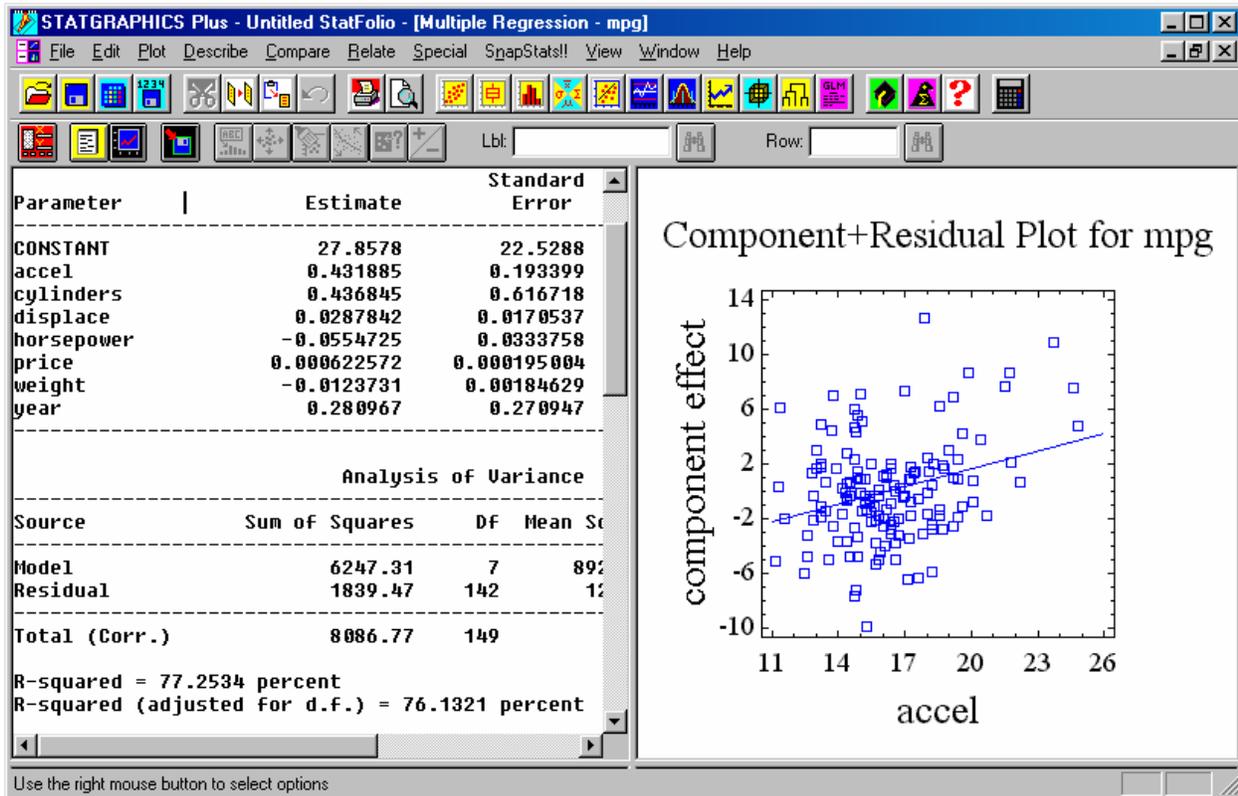
Example of means report and means plot in One-way ANOVA:



8. THE MULTIPLE REGRESSION PROCEDURE

The **multiple regression** procedure is used to estimate linear relationships between variables. For example, can mpg be predicted as a linear function of other variables such as weight, horsepower, etc.? To find out, select "Relate/Multiple Regression" from the main menu. On the Data Input panel, specify "mpg" as the dependent variable, and then specify a number of the other variables as independent variables—e.g., acceleration, weight, horsepower, displace, price, cylinders, and perhaps also year. (It would not make much sense to use "origin" as an independent variable here, because it is not coded in a way that makes a linear relationship plausible. But there are other things we could do with "origin" that we'll get to in a minute. We also cannot use "make" as an independent variable because it is not numeric.) After clicking OK, you will see the Analysis Summary report, which shows the estimated coefficients and other summary statistics of the regression such as R-squared. You will also see a "Component + Residual" plot for the first independent variable, showing the estimated effect of this variable, as shown in the picture on the following page. Many other plots are available, of course.

If you now *maximize the Analysis Summary report* (by double-clicking it) and look at the t-statistics and p-values, you may notice that several of the independent variables are not statistically significant in this regression (i.e., their p-values are greater than 0.05).



If you scroll down to the bottom of this report, you will find the StatAdvisor report, which points out this fact and suggests removing the least significant variable(s) from the model. You could remove the insignificant variables by hitting the 1st icon on the toolbar to return to the Data Input panel, then deleting them from the list of independent variables. But wait, there's an easier way. While the regression summary report is still maximized, click the right mouse button to view the procedure options. You will notice that the options include forward and backward **stepwise regression**—i.e., the automatic addition or deletion of variables from the model based on an F-statistic threshold. (The F-statistic for a variable is the square of the corresponding t-statistics, so including all variables with F-statistics greater than 4 is equivalent to including all variables with t-statistics greater than 2, which are roughly those variables whose estimated coefficients are significant at the $p < 0.05$ level.) Click the "Backwards stepwise" option and then click OK. The regression will be re-run in backward stepwise fashion, and only those variables which exceed the F threshold will be included. If you scroll down to the bottom of the regression summary report, you'll see the details of the stepwise elimination of variables. If you're curious, you can even run it again in forward stepwise fashion and see if you get a different model. (This is not an endorsement of mindless stepwise regression, by the way, but merely a demonstration of some of the tools that are available in Statgraphics. See the caveats below concerning the dangers of automatic selection of forecasting models.)

At this point you may wish to view the full assortment of tables and graphs available for the multiple regression procedure. Click the 2nd and 3rd icons on the toolbar, as usual, and select "All" of the available tables and graphs. The graphs include a number of different views of the residuals and predicted values. The last graph in the series is called an "Interval Plot" for

reasons which may at first be obscure. Maximize this graph by double-clicking on it and look at it closely: this is a plot of the **regression line** superimposed on the data. The default X-axis variable for this plot is the row index, which would be the time index if the data set were a time series. To change the X-axis variable and get some other options, click the right mouse button as usual. You will now see a dialog box offering a selection of other X-axis variables as well as a number of options for plotting **confidence intervals** around the regression line (hence “interval plot”). If you choose to plot confidence intervals for **forecasts** (only), you will observe that a forecast has been made for mpg in row 117: the observation of mpg in this row is missing, while all the other variables are present. The text report called “Reports” shows the numeric values of the forecast and its confidence interval. *This is how you forecast from a multiple regression model in Statgraphics: include rows in the data set where the independent variables are present, but the dependent variable is missing.* For example, if you want to generate forecasts for yet-unobserved cars, you can do so as follows. First, minimize the entire regression analysis by clicking on the down arrow in the top right corner of the “Multiple “Regression” title bar. (Also minimize any other analyses which may be lurking behind the regression analysis.) Then find the “cardata” spreadsheet icon at the bottom of the desktop and double-click on it to bring the data to the screen. Scroll to the bottom of the spreadsheet and enter data for the other variables (except mpg) in one or more new rows. Finally, minimize the spreadsheet and return to the regression analysis: if all independent variables other than mpg are present in any rows you added to the file, forecasts will be automatically generated for them. (Note: to get forecasts for particular data rows when using stepwise regression, you must have “present” observations in those rows for *all independent variables specified on the Data Input panel*, not merely those variables which are included in the final model.)

Some other variations you can now try: return to the Data Input panel (by clicking the first toolbar icon as usual), type **origin=1** in the “Select” box, and click OK to re-run the model. The logical expression **origin=1** creates an indicator (0-1) variable for the cases in which “origin” is equal to 1, i.e., a variable whose value is 1 for every American car and 0 otherwise. (Similarly, **origin=2** would create an indicator variable for the event that “origin” equals 2, i.e., European cars.) Using this variable as the “Select” criterion will restrict the regression analysis to only the American cars. You might be interested to know whether the regression model you have identified for predicting mpg from physical parameters of the car is equally valid for cars made on all three continents. (I found that the models obtained by stepwise regression for individual continents had fewer significant variables than the model for all continents together.) Alternatively, rather than using **origin=1** as a data selection criterion, you could use it as a **dummy variable** in the regression model, in order to test for the presence of an “America effect” that is not explained by the other variables. (Remember that you can type a mathematical expression like **origin=1**, rather than a variable name, on the data input panel for any procedure.) Try this now: *erase* **origin=1** from the “Select” box, and type it instead in the “Independent Variables” box, on the line below the name of the last variable previously used. (Or insert it on another line in the middle of the other variables—the order of entry doesn’t matter.) Click OK to re-run the model. On the Analysis Summary report, look at the *coefficient* of **origin=1** and its p-value. Is it statistically significant? If so, its coefficient represents the amount that should be added to the forecast of mpg when the car is made in America, other factors remaining equal.

One diagnostic plot which should be provided in the regression procedure, but isn't, is a **normal probability plot** to test the assumption of normally distributed errors. It's not hard to get a normal probability of the residuals, though. First, click on the 4th icon on the Analysis Window toolbar to save output variables to spreadsheet. A dialog box will appear listed the various regression outputs that can be saved—click on the “Residuals” item to select the residuals for saving, then hit the OK button. The residuals will now be saved in a column on the spreadsheet. Next select “Plot/Exploratory Plot/Probability Plot” from the main menu, and specify “residuals” as the variable to plot. The normal probability plot will then be generated as a separate analysis. When you wish to return to the regression analysis, just minimize the normal probability plot. If you run further regressions and want probability plots of their residuals, just save their residuals under the same name (i.e., “residuals”) and return to the normal probability plot analysis you previously minimized. It will now be redrawn with the new residuals. In a similar fashion, if you also want to analyze the **autocorrelations** and other time series properties of the saved residuals, you can do so by going to the “Special/Time Series Analysis/Descriptive Methods” procedure. This provides a full range of time series plots, autocorrelation and partial autocorrelation plots, and frequency spectrum plots, as well as a corresponding set of tabular reports. (The Statgallery example on page 2 includes both a residual probability plot and residual autocorrelation plot.)

The **Advanced Regression** module (under “Special” on the main menu) provides probability and autocorrelation plots and many other useful regression options (e.g., General Linear Models, all-subsets regression, ridge regression, nonlinear regression, logistic and probit regression, and Poisson regression, all with out-of-sample validation).

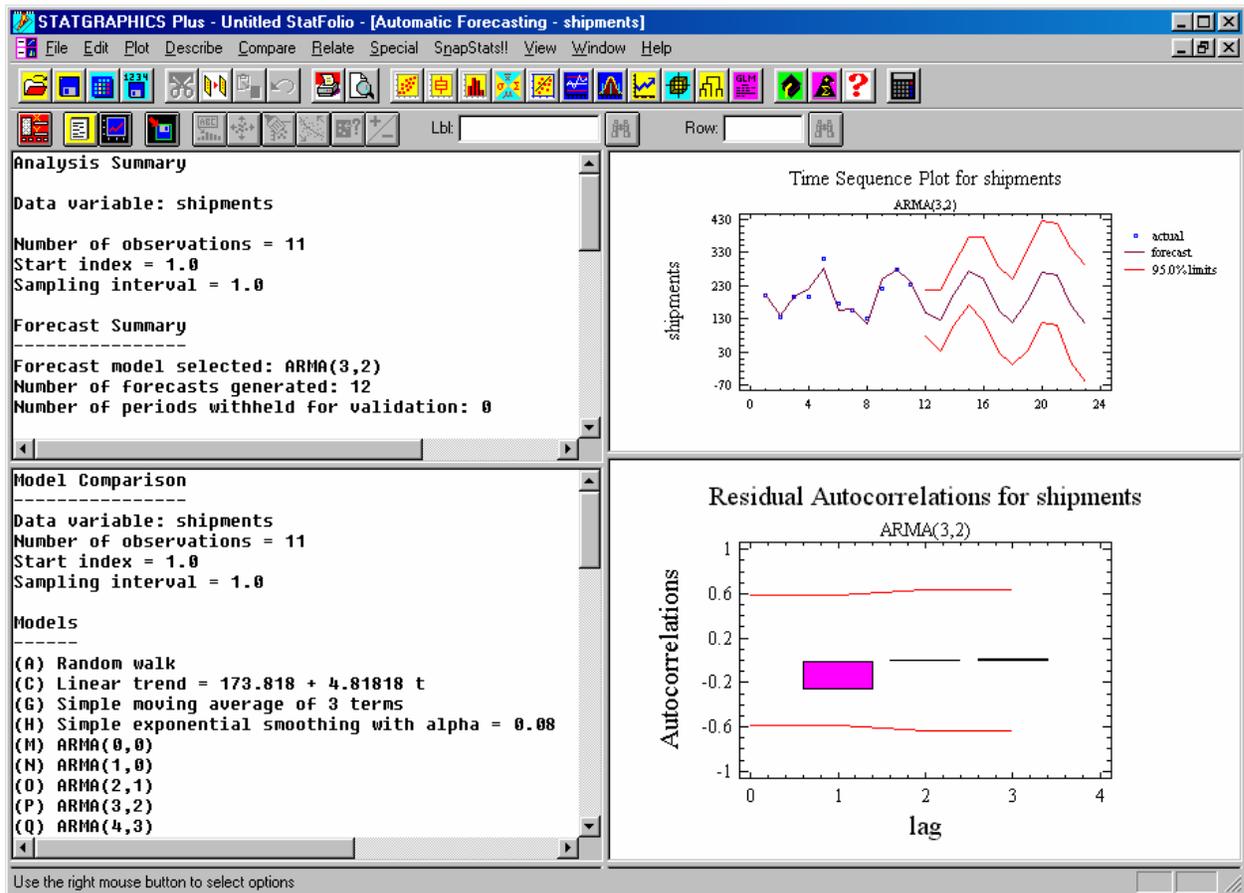
9. SNAPSTATS & AUTOMATIC FORECASTING

Statgraphics 5 includes a menu option called **Snapstats!**, which provides a shortcut to commonly-used procedures such as one-sample analysis, two-sample analysis, etc. Unlike regular Statgraphics procedures, which give you menus of many reports and graphs to be displayed in separate windows, each of which can be separately customized, the Snapstats procedures give you an instant array of reports and graphs on a single page of output. This could be handy for doing a quick exploratory analysis of your data, although the modeling and reporting options are more limited than in the full-blown versions of the same procedures found elsewhere on the Statgraphics menu.

One of the more intriguing Snapstats procedures is **Automatic Forecasting**: plug in a time series, out comes a forecast! If you are tempted to try automatic forecasting in Statgraphics, I recommend that you *don't* use the Automatic Forecasting procedure on the Snapstats menu. Instead, use the new Automatic Forecasting procedure on the Special/Time Series Analysis menu. The latter procedure is the same EXCEPT that it provides you with more control over the model-fitting process and much more complete output, such as the table of comparisons between different models that were tested and detailed reports and charts of the forecasts and residuals. Like automatic forecasting procedures in other statistical packages, the AF procedure in Statgraphics works by conducting an internal tournament among members of several classes of commonly used models: trend models (flat, linear, quadratic, exponential, S-curve), smoothing

models (random walk; simple, linear, and quadratic exponential smoothing; and Winters' seasonal exponential smoothing), and ARIMA models. The tournament winner is determined by ranking the models according to an "information criterion" such as the Akaike or Schwarz criterion, which (roughly speaking) is a measure of the root-mean-square error of the model adjusted for the number of coefficients estimated, with a heavier penalty for model complexity than is provided by the usual RMSE statistic. (Note: ARIMA stands for "Auto-Regressive Integrated Moving Average." The Automatic Forecasting procedure in Statgraphics actually uses a restricted class of ARIMA models called ARMA($n,n-1$) models, in which the number of AR coefficients is constrained to be one more than the number of MA coefficients, and the data are not stationarized by differencing before estimating the coefficients.) These types of models and ranking criteria are covered in my elective course on forecasting and in popular textbooks on business and economic forecasting. If you don't know anything about them, it might be rather dangerous to use this procedure indiscriminately. Automatic model-selection methods do not absolve you of responsibility for understanding your own data and models, and they carry a substantial risk of over- or under-fitting the data in spectacular fashion. For example, if you put in a short time series that has no obvious pattern, the AF procedure may end up selecting a model that produces a beautiful—and completely nonsensical—cyclical pattern of forecasts, like this:

Example of automatic forecasting (caveat emptor!)



Statgraphics also contains an excellent “manual” forecasting procedure under **Special/Time Series Analysis/Forecasting** on the main menu. The latter procedure allows you to fit all of the same models that are used in the Automatic Forecasting procedure, plus many others, including time series models with regressors (independent variables) and ARIMA models with differencing and arbitrary numbers of AR and MA coefficients. The key to using the manual Forecasting procedure is to use the *right mouse button* to select Analysis Options, where you can specify up to *five different models* to be tested simultaneously. (They are assigned code letters ABCDE, selected by radio buttons.) In any case, if you want to know the *coefficients* of the model that was fitted in the Automatic Forecasting procedure, you may have to fit the same model again in the manual Forecasting procedure, because the Analysis Summary report in the automatic version does not show the coefficients.

10. ON-LINE DOCUMENTATION

Statgraphics comes with its own extensive on-line documentation and help, including the “StatWizard” and “StatAdvisor” features described above, plus a standard Windows help file and a set of detailed procedure manuals and tutorials in pdf form. The help files and pdf manuals can be accessed from the “Help” option on the main menu. Alternatively, you can find the pdf files for the manuals and tutorials on your hard disk in the directories called (usually):

C:/Program Files/STATGRAPHICS PLUS 5.0/Manuals and **C:/Program Files/STATGRAPHICS PLUS 5.0/Manuals/Tutorials**. There is also a Statgraphics web site at (naturally) <http://www.statgraphics.com>.

* * *

This brief tutorial has not been intended to provide an exhaustive overview of all the features available in Statgraphics, nor a demonstration of perfectly correct data analysis, but it should give you a feel for the kinds of analysis that are possible. Now that you know how to use the program interface and some of the commonly used procedures, you should be able to learn your way around the rest of the program by simply exploring the menu, checking out the right-mouse-button options, using the Analysis Window toolbar to generate additional tables and graphs, and by using the on-line help system and pdf manuals.