

SIMSTAT

for Windows

User's Guide

Designed and written by Normand Péladeau

Provalis Research

DISCLAIMER

This software and the disk on which it is contained are licensed to you, for your own use. This is copyrighted software owned by Normand Péladeau. By purchasing this software, you are not obtaining title to the software or any copyright rights. You may not sublicense, rent, lease, convey, modify, translate, convert to another programming language, decompile, or disassemble the software for any purpose. You may make as many copies of this software as you need for backup purposes. You may use this software on more than one computer, provided there is no chance it will be used simultaneously on more than one computer. If you need to use the software on more than one computer simultaneously, please contact us for information about site licenses.

WARRANTY

The SIMSTAT product is licensed "as is" without any warranty of merchantability or fitness for a particular purpose, performance, or otherwise. All warranties are expressly disclaimed. By using the SIMSTAT product, you agree that neither Normand Péladeau nor anyone else who has been involved in the creation, production, or delivery of this software shall be liable to you or any third party for any use of (or inability to use) or performance of this product or for any indirect, consequential, or incidental damages whatsoever, whether based on contract, tort, or otherwise even if we are notified of such possibility in advance. (Some states do not allow the exclusion or limitation of incidental or consequential damages, so the foregoing limitation may not apply to you). In no event shall Normand Péladeau's liability for any damages ever exceed the price paid for the license to use the software, regardless of the form of claim. This agreement shall be governed by the laws of the province of Quebec (Canada) and shall inure to the benefit of Normand Péladeau and any successors, administrators, heirs, and assigns. Any action or proceeding brought by either party against the other arising out of or related to this agreement shall be brought only in a PROVINCIAL or FEDERAL COURT of competent jurisdiction located in Montréal, Québec. The parties hereby consent to in personam jurisdiction of said courts.

COPYRIGHT

Copyright © 1996 Normand Péladeau. All rights reserved. No part of this publication may be reproduced or distributed without the prior written permission of Normand Péladeau, 2414 Bennett Street, Montreal, QC, CANADA, H1V 3S4.

Trademarks

IBM-PC and PC-DOS are registered trademarks of International Business Machines Corporation.

Microsoft Windows and MS-DOS are registered trademarks of Microsoft Corporation.

Excel is a product of Microsoft Corporation

SPSS/PC+ and SPSS for Windows are a registered trademark of SPSS Inc.

dBase and Paradox are registered trademarks of Borland International.

Quattro Pro is a registered trademark of Corel Corporation

Lotus 1-2-3 and Symphony are registered trademarks of Lotus Development Corporation.

Other product names mentioned in this manual may be trademarks or registered trademarks of their respective companies and are hereby acknowledged.

Acknowledgments

Special thanks to Marc Aras, Jean Bélanger, Jacques P. Beaugrand, DeWitt Kay, Warren L. Kovach, Mark Thomas Lindemann, Ian D. Livingstone, Rashid Nassar, Ben Riga, Roel van Schaik, George Schwartz, Mark Von Tress, Todd Woodward, and several others for their invaluable support, comment, feedback, and advice during the development of this program.

TABLE OF CONTENTS

Introduction	1
Using this manual	1
Manual conventions	2
1- Getting started	3
The SIMSTAT package	3
System requirements	3
Installing SIMSTAT	3
Making a backup copy	3
Starting the program	4
Startup options	4
The working environment	4
Changing the active window	6
Working with pull-down menus	6
Working with dialog boxes	6
Toolbar	9
Status bar	11
Getting help	11
2 - Tutorial: Performing statistical analyses	13
Step 1 - Opening a data file for analysis	13
Step 2 - Assigning variables	13
Step 3 - Performing frequency analyses	15
Step 4 - Viewing the results	16
Step 5 - Selecting a subset of cases	17
Step 6 - Performing regression analyses	18
Step 7 - Keeping hard-copies of results	19
Step 8 - Exiting SIMSTAT	19
3 - Data file operations	20
Opening an existing data file	21
SIMSTAT for Windows data files	22
Navigating in the Data Window	23
Creating a new data file	24
Defining variables	27
Entering and Editing Data	31
Filtering records	33
Sorting records	36
Importing data from other applications	38
Exporting data to other applications	40
Merging data files	42
Limiting access to data files	43
Archival backup of data files	45
Creating and using variable sets	46

4 - Data transformation	48
Examining data distributions	49
Computing values	50
Performing conditional transformations	55
Quick transformations	56
Recoding values of a variable	57
Transforming a variable into ranks	58
Dummy recoding of nominal variables	58
Numeric recoding	59
5 - Working with the notebook	60
Navigating in the notebook	61
Using the notebook index	62
Output management using Tabs	63
Rounding numerical values	64
6 - Statistical analysis	65
Assigning variables for statistical analysis	65
Binomial test	67
Bootstrap analysis	68
Full analysis bootstrap analysis	73
Breakdown analysis	74
Correlations	75
Crosstabs	77
Descriptives	80
Factor analysis	81
Frequencies	86
Friedman Test	89
GLM ANOVA/ANCOVA	90
Inter-raters analysis	97
Item Analysis	100
Kolmogorov-Smirnov 1 Sample Test	108
Kolmogorov-Smirnov 2 Samples Test	109
Kruskal-Wallis	110
Listing cases	111
Logistic regression	112
Mann-Whitney U test	114
McNemar test	115
Median test	116
Moses test of extreme reactions	117
Multiple regression	118
Multiple responses analyses	125
Nonparametric matrix	127
One sample chi-square test	129
Oneway ANOVA	130
Regression analysis	133
Reliability analysis	136
Runs test	140
Sensitivity analysis	141

Sign test	143
Single-case design	144
Time-series analysis	146
T-test analysis	149
Wilcoxon test	152
7 - Working with charts	153
Creating specific charts	153
Navigating in the chart window	155
Customizing charts	158
Editing titles and axis labels	158
Editing axis scaling and grid	159
Controlling the legend display	160
Displaying data point values	160
Modifying specific chart options	160
Setting global options	160
3D View	162
Zooming in and out	163
Exporting charts	165
8 - Using scripts	167
Introduction to the scripting feature	168
Using normal and encrypted script files	168
Opening an existing script	169
Navigating in the script window	170
Using the RECORD SCRIPT Feature	171
Running a script	171
Saving script files	172
9- Script language reference	173
Syntax Convention	173
Using memory and data file variables	174
Expression Operators and Functions	176
One line descriptions of commands	178
Commands description	181
10 - Customizing the tools menu	233
11 - Setting the program preferences	235
APPENDIX A - xBase syntax and functions	240
APPENDIX B - References	249
APPENDIX C - Limitations	253
APPENDIX D - Technical support	254

INTRODUCTION

Welcome to SIMSTAT for Windows. This program is designed to provide an easy to use, yet powerful statistical package for scientific, business and engineering applications as well as for teaching purposes. It provides unique features that facilitate data analysis as well as other tasks such as data preparation, output management, and data presentation.

SIMSTAT provides a wide range of statistics including summary statistics, crosstabulation, inter-rater agreement statistics, frequency and breakdown analysis, n-way analysis of variance and covariance, paired and independent t-tests, linear, nonlinear, and multiple regression analysis, time-series analysis, and many nonparametric analyses. Also, SIMSTAT provides powerful bootstrap simulation analyses. These simulations are based on resampling procedures and can be used to provide nonparametric estimates of sampling distributions, to assess the stability of multivariate solutions or to perform nonparametric power analysis.

The program also includes a powerful script language that allows automation of statistical analyses and creation of interactive tutorials, demonstration programs with multimedia features, and even computer assisted testing or interviewing programs.

Using this manual

This user's manual provides all the information you'll need to install and use SIMSTAT. The manual assumes that you are familiar with the basics of operating an IBM PC (or compatible) under Windows 3.1 or later. If you are unfamiliar with mouse actions such as double-clicking and drag and drop operations, or if you're not familiar with windows management, consult your Windows user's manual. This manual is not intended to teach statistics, and assumes that you have sufficient knowledge of statistics to choose the analysis appropriate for your needs. The manual is divided into the following sections:

Introduction - The first part provides detailed instructions on how to install SIMSTAT on your computer and gives a variety of information about the program.

Getting started - The Getting Started chapter provides basic information about the program's operation.

Tutorial - This chapter includes a short tutorial that will steer you through an easy-to-follow example. The tutorial is designed to quickly familiarize you with the basic operation of the program. If you are already familiar with the operation of SIMSTAT, you may skip the tutorial section.

Data file operation - This section introduces you to various tasks that may be performed on data files such as how to open an existing data file, create a new one, import a file from another application, filter records, or sort them on one or several variables. This section also presents instructions on how to perform basic data entry and data editing as well as a description of several security features including file archiving and file access protection.

Data transformations - This chapter presents features to perform transformation of existing variables or to create new variables from computation on existing ones.

Working with the notebook - The notebook window provides an efficient way to browse and manage outputs. This section provides basic instructions for navigating the notebook, editing its contents, and using tabs and the index to manage the outputs.

Statistical Analysis - This chapter provides a description of every statistical analysis appearing under the STATISTICS menu. To allow you to quickly find the information you need, the statistical commands are presented in alphabetical order.

Working with charts - This section describes of the steps involved in the creation, modification, printing and saving of charts.

Using scripts - This section introduces you to various tasks that may be performed with script files such as how to open an existing script file, how to execute it, or how to use the recoding feature to easily create script.

Script language reference - This section outlines the syntax conventions of the various commands and options, and provides an alphabetical listing of all script commands.

Customizing the TOOLS menu - This section provides instructions on how to add programs to, delete programs from, or edit programs using the TOOLS menu.

Setting the program preferences - This section provides a description of all global options affecting the program's working environment, file handling, and printing.

Appendices - The last part of this manual contains 4 appendices that deal with the following topics:

Appendix A - Description of xBase syntax rules used in the FILTER, SORTING and COMPUTE commands and of each xBase function.

Appendix B - Various suggestions for further reading on statistics.

Appendix C - Program limitations.

Appendix D - Obtaining technical support.

Manual conventions

The following conventions are used throughout the manual:

- `MONOSPACE` is used to indicate what is to be typed on the keyboard.
- **bold** text is used to highlight keywords, and identify text displayed by the program.
- Words between the less than (<) and the greater than (>) characters indicate a key to be pressed. For example, <Enter> mean you have to press the **Enter** key.
- A dash (-) between key names means you have to press two keys simultaneously. For example, <Ctrl-x> means that you have to press the <Ctrl> key and hold it down while you press the **x** key.

1- GETTING STARTED

In this chapter, we'll get you started using SIMSTAT by showing how to install the program and how to perform some basic operations. This chapter also includes a short tutorial that will guide you through an easy-to-follow example of an analysis. This tutorial will allow you to quickly become familiar with the steps required to perform statistical analyses. If you are already familiar with the operation of SIMSTAT, you may skip this section.

The SIMSTAT package

Your SIMSTAT package includes the following:

- the SIMSTAT user's manual
- two 3 ½ inch program disks
- the Provalis Research license agreement and registration card (send this in to become a registered owner and receive technical support and information about upgrades and related products).

System requirements

SIMSTAT requires a computer running Windows 3.1 or later, 4MB or more of available RAM, and 3.5 MB of hard disk space. A mouse is optional, but highly recommended.

The program does not need a numeric coprocessor but will use it if available. However, a coprocessor is strongly recommended for analysis of large samples or for extensive bootstrap resampling analysis.

Installing SIMSTAT

This manual assumes that your hard disk is drive C: and that the installation disk is on drive A:. You can change the drive and/or directory by making the appropriate substitutions in these instructions.

To install the program on your hard disk:

- Insert Program Disk #1 in the proper diskette drive.
- If Windows is not currently running, start Windows by typing `WIN` and press **<Enter>**.
- From the Program Manager, choose `RUN` from the `FILE` menu.
- Type `A:SETUP` in the command line box and choose **OK**.
- Follow the prompts on screen to complete the installation.

Making a backup copy

Be sure to make a backup copy of the distribution disks, in case you lose or damage the originals. Keep the original disks in a safe place, away from direct heat, dust, and magnetic sources.

Basic program information

Starting the program

To run SIMSTAT for Windows, start Windows if necessary. If you run Windows 3.1, display the program manager icon, double-click the SIMSTAT for Windows group icon to open it, and double-click the SIMSTAT program icon.

If you run Windows '95, click on the START button, then point to Programs, point to the SIMSTAT for Windows item and then the SIMSTAT for Windows program icon.

Startup options

You can edit the program properties dialog box to add a script filename to the SIMSTAT's parameter list. When you do this, SIMSTAT will automatically execute the script when it starts up. By creating shortcuts (or icons) of the program with different script file names as parameter, it becomes possible to execute those scripts from the desktop with a single mouse click.

The working environment

When you call up SIMSTAT, the first screen you see is the introductory splash screen with product-version information. This screen disappears and brings you to the working environment (see below), which is divided into four parts:

ool bar

Menu bar



Status panel

Working space

- 1) The main menu bar at the top of the screen gives access to nine pull-down menus from which commands can be evoked.
- 2) The tool bar provides quick mouse access to frequently used commands.
- 3) The status panel, located at the bottom of the screen, displays various information about the program's operation.
- 4) The central part of the screen is the working space.

SIMSTAT's working space contains four windows:

Data window - The data window is a spreadsheet like data editor where data values can be entered, browsed, and edited. Each column contains data for one variable, and each row contains the responses for a single case. You can have only one Data window open at a time. The associated Data pull-down menu offers a wide range of procedures to select a subsample of cases, sort cases, transform existing variables, or to create new variables from transformation of existing variables.

Notebook window - The Notebook window displays the statistical output for all analyses performed during a session. The notebook metaphor provides an efficient way to browse and manage output. The text output of each analysis is displayed on a separate page. You can turn pages either with the mouse (by clicking on the page-flip icons at the bottom of the notebook), or by using the <PgUp> and <PgDn> keyboard keys. Tabs may also be added to create sections in the notebook, allowing you to store different types of output in different sections of the notebook. An index of all analyses included in the notebook is also available. This index can be used to quickly locate and go to a specific page, move pages within the notebook, or delete some pages. While each page can be annotated or edited, it is also possible to add empty pages for recording ideas or remarks, sketching an analysis plan, or interpreting results.

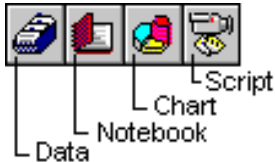
Chart window - The Chart window displays all high-resolution charts created during a session. The associated CHART menu can be used to modify almost any chart feature, such as the axes, the labels, or the legends. You may also reorder the charts, save those charts on disk, print them, or export them either to disk or to the clipboard in Windows Metafile, Windows bitmap, or tab-separated values formats.

Script Window - The script window is used to enter and edit SIMSTAT commands. These commands can either be read from a script file on disk, typed in by the user, or automatically generated by the program. When used with the RECORD feature, the script can also be used as a log window to keep track of the analyses performed during a session. These commands may then be executed again, providing an efficient way to automate statistical analyses. Additional commands also allow one to create demonstration programs, computer-assisted teaching lessons, and even computer-assisted data-entry programs.

Changing the active window

To change the active window, you can use one of three techniques:

- Click on a visible part of another window.
- Choose another window from the WINDOWS menu.
- Click on the proper icon on the toolbar (see below).



Working with pull-down menus

To select a menu item:

- Point to the menu name and click the left mouse button, or press the <Alt> key and the underlined letter in the menu name.
- Point to the command you want to execute and click the left mouse button again, or press the underlined letter in the command you want.

Once in a main menu, select a sub-menu command by typing the underlined letter of the command name. You can also use the <Up> and <Dn> arrow keys to highlight the command you want and press <Enter> to select your choice. Use the <Right> or <Left> arrow keys to move from one pull-down menu to another.

Press <Esc> or click anywhere outside the pull-down menu to close it.

When a menu item is dimmed, the command is currently unavailable. For example, when no data file has been opened, several commands of the EDIT, DATA, and STATISTICS menu are dimmed.

Some menu items also display a keyboard key or key combination to the right of the command name. When such a keyboard shortcut exists, you can bypass the menu system and choose those commands by pressing their associated keyboard shortcuts.

Working with dialog boxes

Some SIMSTAT commands in submenus are followed by an ellipsis (three dots), which indicates that a dialog box appears when you choose this command. Dialog boxes allow you to specify various options used to adjust the program's operation, to customize the analysis output, or to specify conditions to be fulfilled by an analysis.

Click on any item to position the cursor on it. You may also use the following keyboard keys to navigate through dialog boxes:

Key	Effect
<Tab>	Moves to the next item
<Shift-Tab>	Moves to the previous item

To move directly to an item, press the <Alt> key and the underlined letter in the associated text.

The action needed to edit the values of the various options depends on the type of input field. Options panels can contain several types of input field:

Edit box - The Edit box is a rectangular box in which you type a value using the keyboard. They are used to enter a string of characters, a number, a filename, etc.. When you are positioned on such a field, a blinking text-cursor appears. The following table lists the keys that can be used while editing a data entry field.

Key	Effect
<Right arrow>	Moves the cursor one character to the right.
<Left arrow>	Moves the cursor one character to the left.
<Ctrl-left>	Moves the cursor to the beginning of the previous word.
<Ctrl-right>	Moves the cursor to the beginning of the next word.
	Deletes the character to the right of the cursor.
<Backspace>	Deletes the character to the left of the cursor.
<Home>	Moves the cursor to the beginning of the line.
<End>	Moves the cursor to end of the line.
<Ins>	Pressing this key toggles between the insert and overwrite modes. In insert mode, characters are inserted at the cursor position, pushing text to the right of the cursor even further right. When in overwrite mode, characters at the cursor position are overwritten.
<Esc>	Aborts the editing process and restores the previous value.

List boxes - A list box allows you to select a string from a list of valid possibilities. A list box can contain a vertical scroll bar. To make a selection, scroll, if necessary, then click the item you want.

Drop-down list boxes - A drop-down list box is similar to a list box, but its items are hidden from view. To select an item from the list, click on the downward pointing arrow on the right side of the box, scroll, if necessary, then click the item you want.

Check boxes - A check box is a small square box with a text description to its right. Check boxes operate independently of one another. You can turn an option ON or OFF by clicking in the box. If an option with a check box is turned on, an "X" appears in the box.

Radio buttons - A radio button is a small round button. Radio buttons are used in groups to present mutually exclusive options. Click the button to turn the option on; to turn it off, select a different radio button. If a radio button option is active, it contains a dark circle.

Spin buttons - Sometimes an edit box will be presented with spin buttons to its right. This indicates that a numeric value is expected. You can type this numeric value using the keyboard or use the spin buttons to quickly increment or decrement the value shown in the edit box.

Multiple-page dialog boxes

Some dialog boxes may contain a fairly large number of options. In order to facilitate the location and setting of those numerous options, multiple-page dialog boxes have been used to functionally group those options on different pages. For example, many statistical analyses contain a separate page for all options related to the production of high resolution charts while the numerical analyses options are located on another page. You can access a particular page in the dialog box by clicking on its associated tab located at the bottom of the dialog box.

After the values have been edited to suit your needs, you have to click on the **OK** button (in some dialog boxes this button is named **APPLY**) to accept those values and proceed. If you want to leave the dialog box, restore the previous values, and suspend the current operation, just click on the **CANCEL** button. A **HELP** button is often displayed to give you access to the context-sensitive help file. Some dialog boxes also provide additional command buttons, which give you access to special functions.

Toolbar

The main toolbar is displayed across the top of the application window, below the menu bar. The toolbar provides quick mouse access to many commands used in SIMSTAT.

To hide or display the toolbar, you can use the Preferences menu command and set the Show Toolbar check box accordingly.

When the mouse cursor rests over a toolbar button for more than one second, a small help hint appears, displaying a short description of this button. To turn these help hints on or off, use the Show Tool Hints option in the Preferences dialog box.

The following list describes the various icons used by SIMSTAT.

Click To

The following eight buttons may be present no matter which window is currently active. However, their action differs, depending on which window is active.



Create a new file. You may use this button to clear the content of the currently active window.



Open an existing file of the same type as the active window. SIMSTAT displays the Open dialog box, in which you can locate and open the desired file.



Save the active window document with its current name. If you have not named the document, SIMSTAT displays the Save As dialog box.



Print one or more pages or charts of the active window.



Cut the selected text (Notebook or Script window) or selected cell (Data window)



Copy the selected text (Notebook or Script window) or selected cell (Data window)



Paste text or a value from the clipboard.



Erase the selected text (Notebook or Script window), selected cell (Data window) or chart. When viewing the notebook index, pressing this button erases the currently selected page.

Notebook window buttons



Add a tab to the Notebook.



Add a blank page to the Notebook.



Select variables to be included in subsequent statistical analysis.

Chart window buttons



Copy a bitmap image of the current chart to the clipboard.



Turn on/off the 3D option for the current chart.



Display the 3D rotation dialog box.



Zoom in the currently displayed graph.

Script window buttons



Run the entire script.

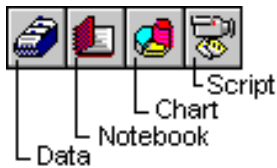


Run the selected commands.



Run the script from the current line.

At the right end of the toolbar, you will see four buttons representing SIMSTAT's four different windows.



Click on one of these icons switch to another SIMSTAT window.

Status bar

The status bar, at the bottom of the screen, displays information about the current session including the number of selected variables, and the amount of free Windows resources and memory. It also displays useful information about the current output page (date and time of creation, data file name) or about the current chart (data and time of creation). The left-most panel contains a gauge that is sometime used to indicate the progression of a task.



Getting help

No matter where you are in SIMSTAT, you can get more information about the task you're working on by pressing <F1>. This function key accesses SIMSTAT's context-sensitive help.

You may also click on the **HELP** button from any dialog box to obtain help on the various options available in the dialog box.

Alternatively, you may use the various commands in the HELP menu to display a table of contents of available help topics or search for a specific topic.

From the help window, you can copy, paste, annotate, and print help text using commands in the FILE and EDIT menu.

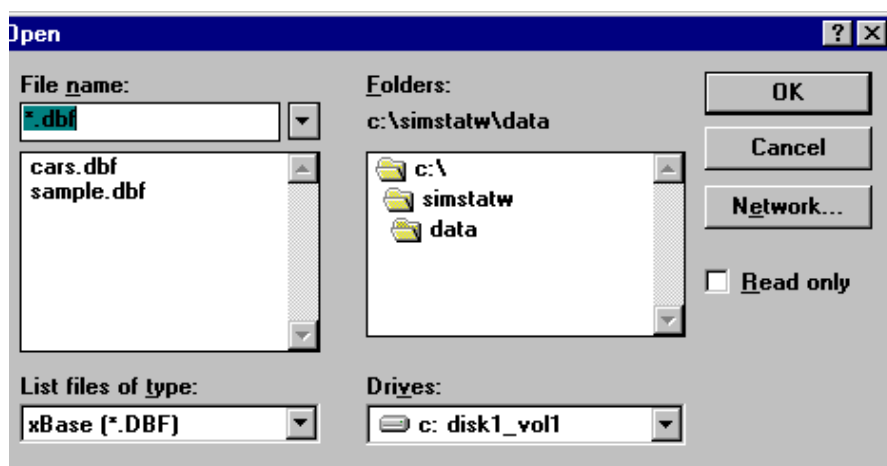
2 - TUTORIAL: PERFORMING STATISTICAL ANALYSES

The following section provides a tutorial for beginning users of SIMSTAT. It contains step-by-step instructions for computing descriptive statistics and performing a regression analysis on a sample data file. The data is from a fictitious study about the influence of various factors on children's aggressive behavior during school recreational activities. This tutorial uses data stored in the dBASE file SAMPLE.DBF that comes with your program disk.

Step 1 - Opening a data file for analysis

The first step to performing a statistical analysis is to open a data file that you want to analyze.

After calling up SIMSTAT, select the DATA | OPEN command from the FILE menu. This command pulls down an Open File dialog box that lists all valid files and subdirectories.



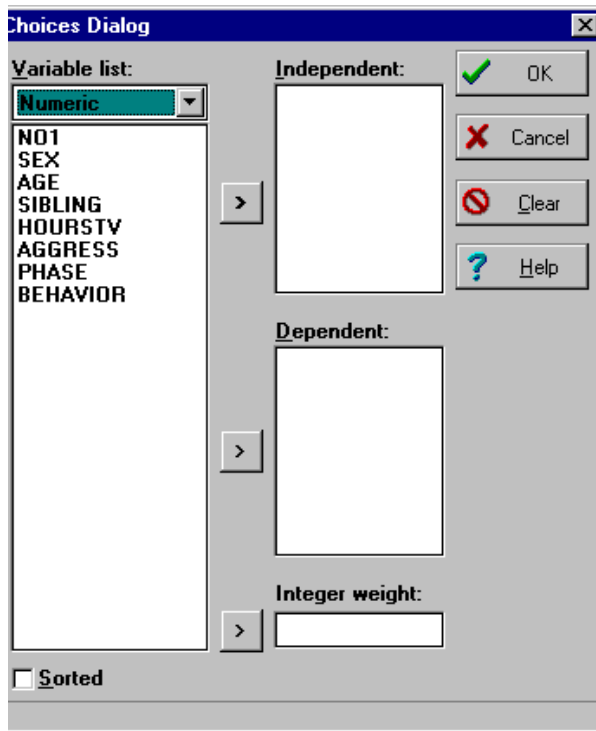
Double click on the SAMPLE.DBF file to open it.

If the file is not on the current drive or in the default directory you may use the following methods to locate the data file:


- If the file is on a different disk, click on the down arrow of the Drives list box to display available drives and select the disk where the file is located.
- If the file is in a different directory, double-click on the directory names in the Directory list box (also called Folder) list box to move through the directory tree.

Step 2 - Assigning variables

The next step is to choose the variable(s) on which you will perform the analysis. Select the CHOOSE X-Y command from the STATISTICS menu to display the Variables Selection dialog box.



This dialog box contains, among other things, 3 different list boxes. The one located to the left of the dialog box contains a list of all variables in the data file. The selection of variables for upcoming analysis is carried out by moving the proper variable's name(s) from this list to the independent or dependent list box. When doing descriptive analyses on separate variables, it does not matter whether a variable is assigned as dependent or independent. For the current example, we will thus move the variables to the independent variable list box. To achieve this:

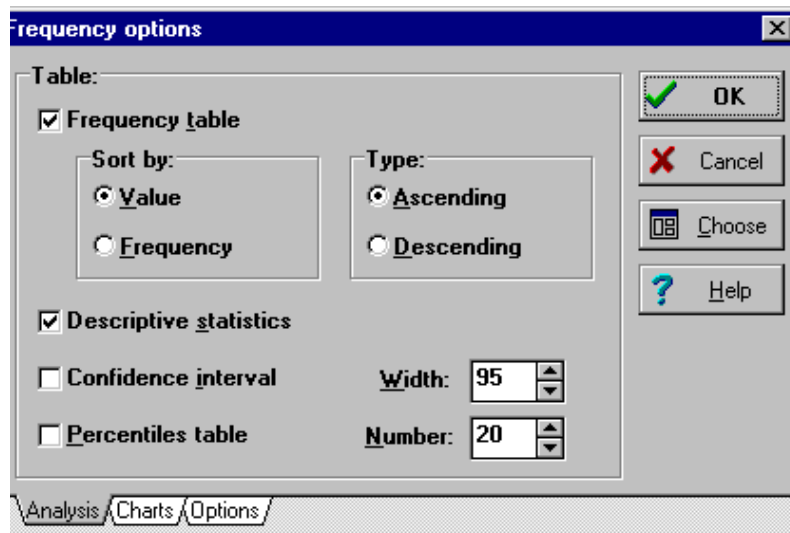
- Highlight the AGE variable (age of the child) in the variable box by clicking once on it.
- Click on the  button just beside the independent list box to move the highlighted variable name to this list box.
- Using the same procedure, move the variable SIBLING to the independent list box.

NOTE: If you select the wrong variable, you can remove the variable from the Independent list box by clicking on the variable name. When a name in the independent list box becomes highlighted, the arrow of the button beside this box changes direction, indicating that pressing this button will put the variable back in the variable list box.

- Click on the **OK** button to close the dialog box and activate the variable selection. If you want to cancel the operation and restore the previous variable assignments, click on the **CANCEL** button.

Step 3 - Performing frequency analyses

To perform a frequency analysis on the variables we have just chosen select DESCRIPTIVE | FREQUENCIES from the STATISTICS menu. The following dialog box appears on screen.



Using this dialog box, we will instruct the program to print, for every selected variable, a frequency table sorted in ascending order of value, detailed descriptive statistics and a histogram. To achieve this:

- Activate the Frequency Table check box (a check will appear in it).
- Set the Sort By option to Value and the Type option to Ascending.
- Activate the Descriptive Statistics check box.
- Deactivate the Confidence Interval and Percentiles Table check box.

The dialog box looks similar to the one shown above.

All graphing options are located in the second page of the dialog box. To activate this page, click on the Charts tab at the bottom of the dialog box. To produce a histogram, make sure that the histogram check box is the only one containing a check mark. If other options are enabled, click on those check boxes to deactivate them.

After setting the proper options, you must activate them and tell SIMSTAT to perform the analysis by selecting the **OK** button. SIMSTAT computes the requested statistics and appends the numerical results to the Notebook window, while the histograms are automatically added to the Chart window.


Step 4 - Viewing the results

To view the numerical output of the analysis just performed, make the notebook window active by either clicking on a visible part of this window, selecting it from the WINDOWS menu, or clicking on the notebook icon at the right end of the toolbar.


The notebook displays the last page which contains the frequency table and descriptive statistics of SIBLING. Use the arrow keys (i.e. <Up> , <Dn>, <Left> and <Right>) to navigate through this page. If you prefer, you may also set the program to display horizontal and vertical scroll bars and use those scroll bars to browse through the active page (see Setting Program Preferences section, page 235).


You may also use the <PgUp> or <PgDn> keys to move one screen at a time. If you press the <PgUp> key while the cursor is already on the top of the page, the program will bring you to the previous page, while pressing the <PgDn> key when the cursor is located at the bottom of a page will move you to the next page (if any).


To move from one page to another, you may also click with the mouse on the page corner icons located at the bottom of the notebook.

 Rather than moving from one button to the other in order to move forward and backward in the notebook pages, you can click on the same button using the right button of the mouse to perform the reverse action.

To view the histograms produced during the same analysis, activate the chart window. The chart window displays the last chart produced.

To display the previous chart in the window press the <Ctrl-P> combination key or click on the  icon.

To display the next chart in the window, press the <Ctrl-N> combination key or click on the  icon.

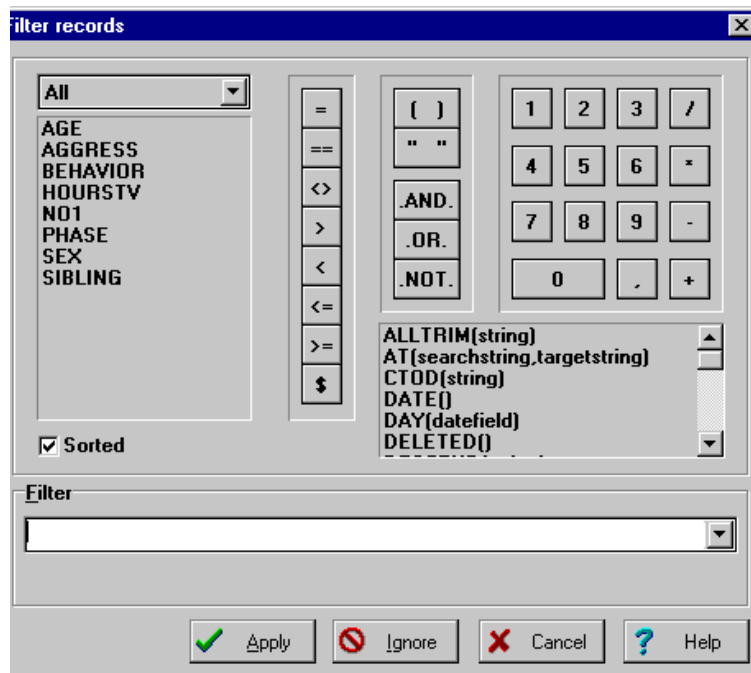
To select a specific chart from a list of all charts, click on the  button to display a list box and select the chart from the list.

When an image is displayed in the Chart windows, it is possible to customize its axis scaling or labels, change the colors and fonts, and modify several other features of the chart. All these customization options can be accessed from the CHART menu. It is also possible to save these charts on disk, export them either to files or to the clipboard, or print them. (For more information on the various chart options available, see the section entitled Working with Charts on page 153).

Step 5 - Selecting a subset of cases

For our second analysis, we will perform two regression analyses where AGGRESS (number of aggressive behaviors) will be used as the dependent variable while AGE and HOURSTV will be treated as two separated independent variables (or predictors). For the purpose of this demonstration, we will restrict the analysis to male subjects.

To select this subgroup, choose the FILTER command from the DATA menu. This command displays a dialog box like the one below.



The Filter edit box allows you to specify a condition that must be met in order to include the case in the analysis. The sex of the child has been stored in a numeric variable named SEX using 1 to designate boys and 2 for girls. To select the boys for the next analysis, type the following condition:

SEX = 1

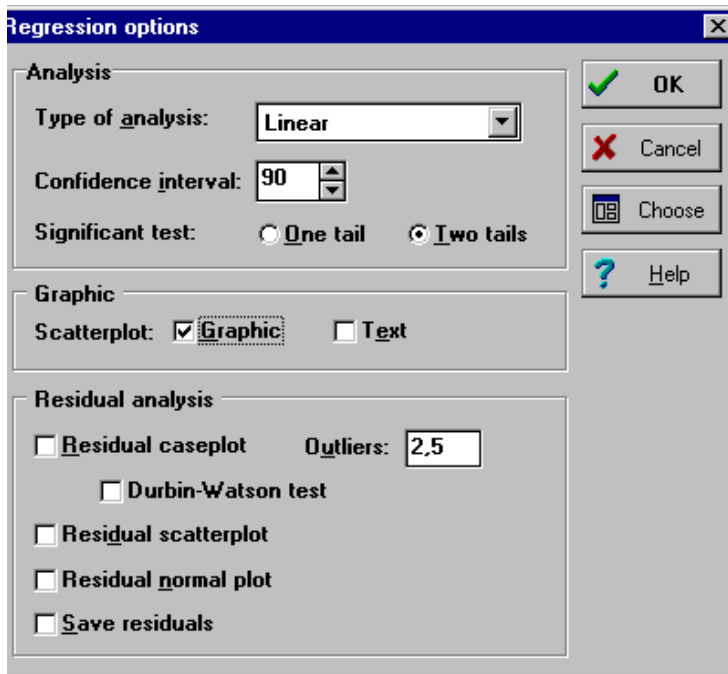
You may also use the filter building buttons located at the top of this dialog box to specify this condition. To build this filtering condition using these buttons:

- Click on the SEX variable in the list box located to the left of the dialog box
- Click on the '=' relational operator button to add an equal sign to the equation.
- Click on the '1' button of the numeric keypad to insert this value after the equal sign.

To exit this dialog box and activate the filtering condition, click on the **APPLY** button.

Step 6 - Performing regression analyses

Choose the REGRESSION | REGRESSION command from the STATISTICS menu. The following dialog box appears:



The first step necessary to perform our regression analysis is to change the currently selected variables. In the previous analysis, the variables were selected prior to the display of the analysis dialog box. For this example, we will alter the variable assignment while editing the regression analysis option. To achieve this, click on the **CHOOSE** button to activate the Variables selection dialog box. Using the previous instructions, assign the AGE and HOURSTV variables to the Independent list box and move the AGGRESS variable to the Dependent list box. Click **OK** to confirm this variable selection and return to the analysis dialog box.

Then, using the proper keys or mouse actions:

- Set the Type of Analysis option to Linear.
- Set the numeric field beside Confidence Interval to 90 in order to obtain a 90% confidence interval on beta weights.
- Select a 2-tailed test by clicking on the proper radio button.

To obtain a scatterplot that will allow you to visualize the relationship between the dependent and independent variables, set the SCATTERPLOT option to Graphic.

Disable the remaining options so that the dialog box looks similar to the one displayed above. When you click on the OK button, SIMSTAT calculates two separate linear regression equations with AGGRESS as the dependent variable and AGE and HOURSTV as predictors. You may browse through the results of this analysis if you so desire.

Step 7 - Keeping hard-copies of results

Once analyses are performed, you will often want to keep copies of the numerical results and charts produced during a session.

To save the results on disk:

- Choose the NOTEBOOK | SAVE command from the FILE menu. A file save dialog box will appear. By default, SIMSTAT use the .SNB extension for notebook files. If no extension is given, the program automatically adds this extension to the end of the file name.
- Enter the name of the file under which you wish to save the notebook and press <Enter> or click on the **OK** button to save the file.
- Choose the CHART | SAVE command from the FILE menu to save all charts in the Chart window in a file name. The default extension for chart files is .CHX. If no extension is given, the program automatically adds this extension to the end of the file name.

To print the numerical results or the charts:

- Activate the Notebook window to print numerical results, or the Chart window to print charts.
- Select the PRINT command from the FILE menu or click on the printer button on the tool bar.
- Set the desired options and click on the **OK** button to start printing.

Note: By default, SIMSTAT is configured to start printing each notebook page at the top of a new page and to print 2 charts per page. To change these options, select the PREFERENCES command from the FILE menu (see section Setting Program Preferences at page 235).

Step 8 - Exiting SIMSTAT

To exit from SIMSTAT, choose the EXIT command from the FILE menu. If, during a session, the contents of any window have been modified, and have not been saved, then dialog boxes will appear, asking you if you would like to save those changes to disk. Choose **Yes** to save the changes to the documents and exit SIMSTAT or choose **No** to continue to exit without saving the documents.

3 - DATA FILE OPERATIONS

The data window is a spreadsheet style data editor where values can be entered, browsed and edited. Each column contains data for one variable and each row contains the responses for a single case. You can have only one Data window open at a time. The associated DATA pull-down menu offers a wide range of procedures to select a subsample of cases, sort these cases, transform existing variables, or create new variables resulting from mathematical operations performed on existing variables.

NAME	MPG	CYLIND	DISPLA	HRSPWR	ACCEL	MODEL
chevrolet chevelle malibu	18,0	8	307,0	130,0	3504,0	12,0
buick skylark 320	15,0	8	350,0	165,0	3693,0	11,5
plymouth satellite	18,0	8	318,0	150,0	3436,0	11,0
amc rebel sst	16,0	8	304,0	150,0	3433,0	12,0
ford torino	17,0	8	302,0	140,0	3449,0	10,5
ford galaxie 500	15,0	8	429,0	198,0	4341,0	10,0
chevrolet impala	14,0	8	454,0	220,0	4354,0	9,0
plymouth fury iii	14,0	8	440,0	215,0	4312,0	8,5
pontiac catalina	14,0	8	455,0	225,0	4425,0	10,0
amc ambassador dpl	15,0	8	390,0	190,0	3850,0	8,5
citroen ds21 pallas		4	133,0	115,0	3090,0	17,5
chevrolet chevelle concours (sw)		8	350,0	165,0	4142,0	11,5
ford torino (sw)		8	351,0	153,0	4034,0	11,0

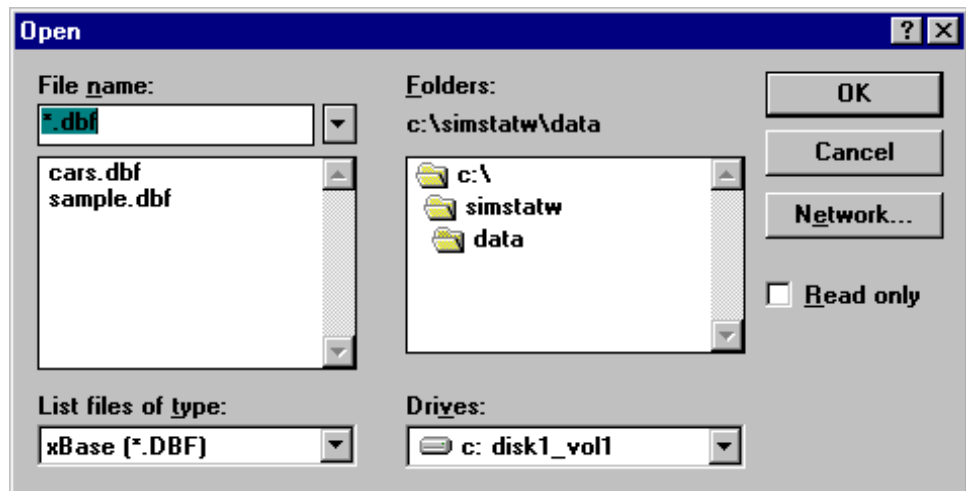
This section introduces you to various tasks that may be performed on data files such as how to:

- Open an existing data file.
- Create a new data file.
- Set the variable appearance and definition (display width, missing values, value labels, etc.).
- Enter and edit values in the data grid.
- Filter records.
- Sort the data grid on one or several variables.
- Import from and export to other applications.
- Save and restore archived copies.
- Restrict the default access to the data file.

Information on data transformation techniques will be introduced in the next chapter.

Opening an existing data file

To open an existing data file, select the DATA | OPEN command from the FILE menu. This opens an open file dialog box as shown below.



When this dialog box is evoked, the program points to the default data directory and displays in the File Name list box all available data files in this directory. To open a file, double click on its name or select it and then click on the **OK** button.

If the name of the data file you want to open is not displayed, type the filename in the File Name box (including drive and path if necessary) and select the **OK** button.

You may also use the following components to locate the data file:

- If the file is on a different disk, click on the down arrow of the Drives list box to display available drives and select the disk where the file is located.
- If the file is in a different directory, double-click on the directory names in the Directory (or Folder) to move through the directory tree.

If the file name is displayed in the File Name list box, double-click on it to open the file, or select it and click on the **OK** button.

If you want to open a data file used previously, click on the down arrow button at the right side of the File Name edit box and select the filename.

SIMSTAT for Windows data files

SIMSTAT for Windows uses the industry standard xBase file format as its own format. Thus, the program can read DBF files created by almost any program that can create these files. SIMSTAT supports dBase files containing up to 1022 fields provided that the maximum record length is less than 64k. SIMSTAT will also create several associated files in order to store important information about the data file. The following list describes the various file extensions that are used:

Extension Description

- *.STR Files with an .STR extension contain all variable information defined by the user such as variable labels, missing values and display format (display width, number of decimal places, alignment). This file also contains information about access restrictions set by the user.
- *.VLB Files with this extension contain all value labels defined by the user.
- *.SET Set files contain information about defined sets of variables.
- *.IDX Files with an .IDX extension are automatically created by SIMSTAT to keep track of the filtering and sorting conditions set by the user.
- *.FPT Files with an .FPT extension are used to store texts entered in memo fields.



When moving data files to a new location, you must also move these related files and place them in the same directory as the DBF data file. To facilitate this task, you may use the ARCHIVES | BACKUP and ARCHIVES | RESTORE commands which will allow you to store all the necessary files in a single archive file and will automatically restore those files in the location of your choice.

Navigating in the Data Window

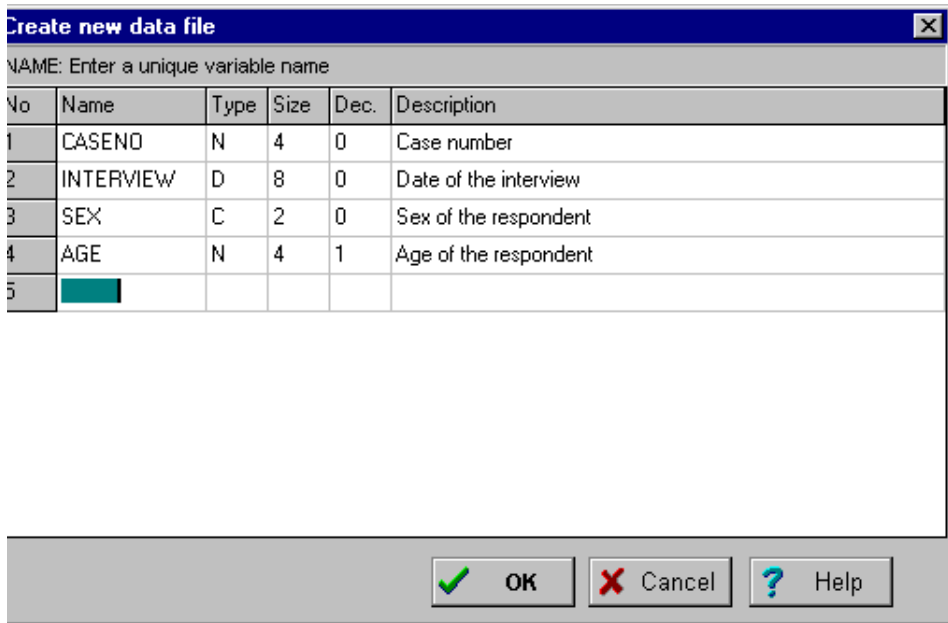
To move the caret to a specific cell, you can click on that cell with the mouse. You can also use the following keys to navigate in the spreadsheet:

<Up>	Move one cell up.
<Down> or <Enter>	Move one cell down. When you reach the last line and hit the down arrow key, a new case is created.
<Right> or <Tab>	Move one cell to the right. When you reach the end of a row, these keys bring you to the first column of the next row. If you reach the last cell and hit one of these keys, a new row is created.
<Left> or <Shift-Tab>	Move one cell to the left. If the cursor is on the first column, pressing either of these keys brings you to the last column of the previous row.
<PgUp>	Move one screen up.
<PgDn>	Move one screen down.
<Home>	Move to the first variable (column).
<End>	Move to the last variable (column).
<Ctrl-Home>	Move to the first record.
<Ctrl-End>	Move to the last record.
<Ctrl-G>	Search for a variable name.

You can also move to a specific value or string within a selected column or anywhere in the spreadsheet by evoking the FIND command in the EDIT menu.

Creating a new data file

To create a new data file, select the **DATA | NEW** command from the **FILE** menu. When this command is evoked, a dialog box similar to this one appears:




No	Name	Type	Size	Dec.	Description
1	CASENO	N	4	0	Case number
2	INTERVIEW	D	8	0	Date of the interview
3	SEX	C	2	0	Sex of the respondent
4	AGE	N	4	1	Age of the respondent
5					

The first step you need to take in order to create a new data file is to define the structure of the file. The File Structure dialog box is a grid entry field type where each row represents a variable in the new data file. This dialog box lets you define various attributes of the new variables such as their name, whether they will contain numeric, alphanumeric values, or dates and their physical width. You can also enter, for each variable, an alphanumeric description up to 60 characters.

Name - The first column of the spreadsheet allows you to enter a name for each variable. Each variable name must be unique (within that data file). Valid variable names begin with a letter and may contain letters, numbers or underscore characters. Punctuation marks, blank spaces, and other special characters are not permitted. The maximum variable name length is 10 characters.

Type - Each variable in the data file must have a type. To specify a variable type, move the cursor to the second column and enter the letter corresponding to the proper data type. SIMSTAT for Windows supports the following types:

Key	Type	Description
C	Character	Character variables can contain up to 254 alphanumeric characters.
N	Numeric	Numeric variables can contain floating point numbers.
D	Date	The date type occupies 8 spaces and holds a year, month and day.
L	Logical	The logical type uses a single space to store a boolean value that can be either true or false (or Yes or No).
M	Memo	The memo type can contain up to 32K of text.

 While SIMSTAT for Windows can perform some analyses such as frequency or crosstabulation on character variables, it is often preferable to use numeric variables, especially when the number of different values of this variable is limited. For example, rather than storing the sex of the respondent as a character variable and using 'male' and 'female' or 'F' and 'M', it is advisable to assign numeric values to this information (for example 1 for male and 2 for female). To facilitate the interpretation of these numeric values, SIMSTAT provides a way of associating an alphanumeric description of up to 60 characters with each numeric value of a variable (see page 28).

Length - The variable length specifies the maximum number of characters that can be stored in the variable. Variable lengths for date and logical field are automatically set to 8 and 1 respectively. The maximum length for a character variable is 254, while the maximum length for a numeric variable is 19.

Decimal - The decimal column lets you define the number of decimal places for numeric variables. Note that the length of a numeric variable with decimals includes the decimal point, a leading zero, and an optional minus sign. The minimum length for a numeric variable that contains one decimal position is therefore 3 (unsigned) or 4 (signed). The maximum number of decimal places permitted by SIMSTAT is 17.

Description - The description column allows the entry of a variable label up to 60 characters long that can be used to describe in more detail the content of the variable. You may leave this column blank if you wish, since it is always possible to later add or edit those labels by using the DEFINE VARIABLE command.

Use the <Up> and <Down> arrow keys to move around the variable list and the <Left> and <Right> arrow keys or the <Tab> and <Shift-Tab> keys to position yourself on the field that you want to modify.

To insert a new variable:

- Position the cursor on the last row and press on the <Down> arrow key.


To modify the position of a variable in the list:

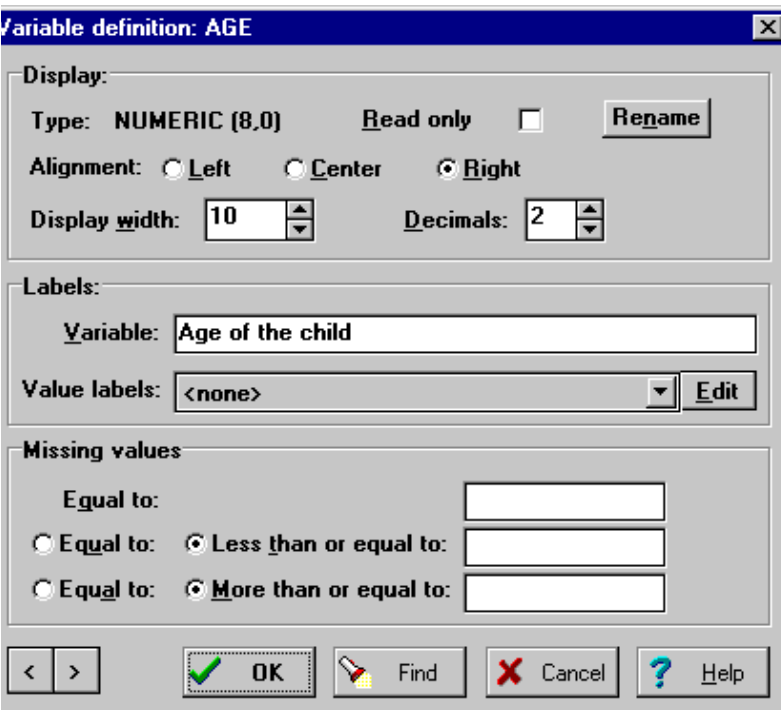
- Click on the left border of the row you want to move.
- While holding down the mouse button, drag the row toward its new location.
- Release the mouse button to drop the variable at its new location.

When you have finished defining the structure information, click on the **OK** button. You will be asked to specify the name of the data file you want to create. If a file with a similar name already exists, you will be asked to confirm the overwriting of this file.

Defining variables

In addition to the specification of the physical structure of the variable, it is also possible to specify a variable's display format (width and number of decimal places), missing values and value labels. To define these attributes use the following steps:

- Position the cursor on the variable you want to define.
- Choose the DEFINE VARIABLE command from the FILE menu or click on the  button on the upper left corner of the data grid. This displays the Variable Definition dialog box as shown below:





The image shows a dialog box titled "Variable definition: AGE". It is divided into three main sections: "Display:", "Labels:", and "Missing values".

- Display:**
 - Type: NUMERIC (8,0)
 - Read only:
 - Alignment: Left, Center, Right
 - Display width: 10
 - Decimals: 2
 - Buttons: Rename
- Labels:**
 - Variable: Age of the child
 - Value labels: <none>
 - Buttons: Edit
- Missing values:**
 - Equal to: []
 - Equal to: Less than or equal to: []
 - Equal to: More than or equal to: []

At the bottom of the dialog box, there are navigation buttons: < >, OK (with a green checkmark), Find (with a magnifying glass), Cancel (with a red X), and Help (with a question mark).

Navigating through variables

The  and  icons located at the bottom of the dialog box can be used to move to the previous or the next variable in the data file.

To locate a specific variable, click on the **Find** button. A dialog box appears that allows you to select a variable name from a list of all variables in the data file. To quickly locate a variable name, you can also type its first letters until the variable name appears in the list box.

Read only - When checked, this option prevents a variable from being modified. This option is useful to prevent accidental or unauthorized changes to the values of a variable. (To prevent the modification of an entire data file, use the DATA | SECURITY command from the FILE menu or open the file as Read Only).

Rename button - You can use this button to change the name of the current variable. When you click this button, you will be asked for a new variable name. This name must not exist in the current data file and should follow the basic rules for valid variable names.

Alignment - The alignment option lets you specify whether the values of this variable should be displayed on the grid aligned to the **left**, at the **center**, or flushed to the **right** of the column. By default, numeric values and dates are flushed to the right of the column while strings are aligned to the left.

Display width - The display width option lets you adjust the display width of the current variable. This option is used exclusively to control how the variable is displayed in the data grid and does not affect the physical size of the variable or its internal precision. While it may be possible to set this option to zero or one character, the actual minimum display width is equal to the number of characters in the variable name. For example, the minimum display width of a variable named AGE will be 3.

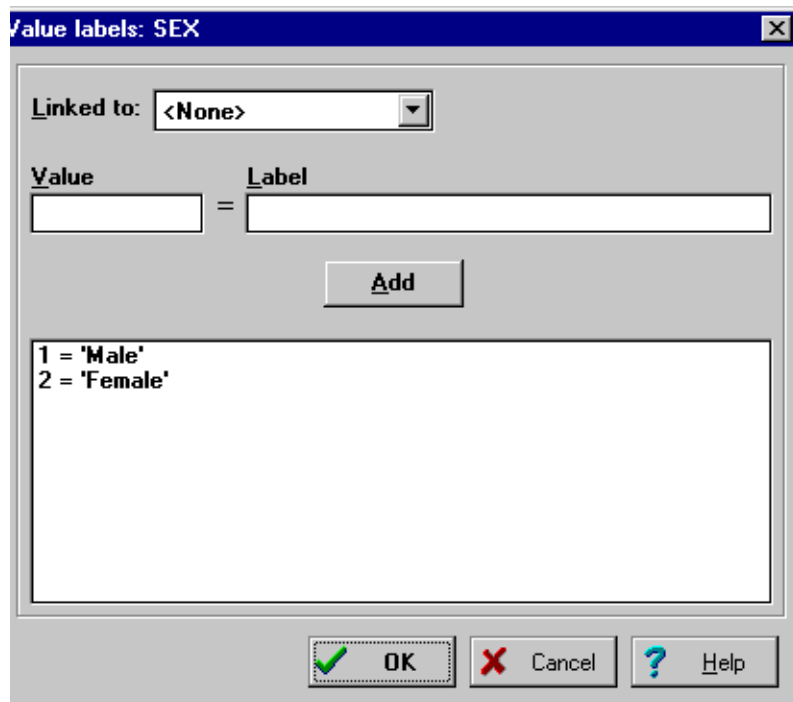
Decimals - When the variable is numeric, the decimals option is used to specify how many decimal places to display in the data grid. This option is used exclusively to control how numeric values are displayed in the data grid and does not affect the internal precision of the variable.

Missing values - In SIMSTAT, any blank cell is treated as a missing value (also called a “system missing value”). However, you may also want to specify the reason for the missing value. For example, in a survey, some respondents may not respond to a specific question because it does not apply to them. They may also refuse to answer, or may have simply forgotten to answer this question. For each numeric variable, it is possible to define up to 3 numeric values that will be treated as missing. When performing statistical analyses or data transformations, all cases containing a blank field or any one of these numeric values will be ignored.

By default, these numeric values are treated as discrete values. However, It is possible to exclude a range of values by treating the second and third missing values as lower or upper limits. For example, if you have a variable containing the ages of respondents, you may choose to exclude all subjects under 18 years by setting the second missing value to 17 (or 17.99 if the variable is measured on a continuous scale with up to 2 decimal places) and selecting the **less than or equal** radio button. You may also exclude cases equal or above a specific value by specifying the upper limit in the third missing value and setting its radio button to **more than or equal**.

Variable label - This option lets you enter an alphanumeric description of the variable of up to 60 characters. SIMSTAT will use this label when displaying statistical results or charts.

Value labels - The value labels feature allows you to assign to each value of a numeric variable a string of up to 60 characters. To define new value labels for the current variable, click on the Edit button next to the value label option. The following dialog box will appear:



To ADD A NEW LABEL:

- Enter a numeric value in the Value edit box.
- Enter the description associated with this numeric value in the Label edit box.
- Press <Enter> or click on the **Add** button to accept this definition.

To DELETE AN EXISTING VALUE LABEL:

- In the value label list box, select the label you want to erase. The caption of the **ADD** button will change to **DELETE**.
- Click on the **DELETE** button

TO MODIFY AN EXISTING LABEL:

- Enter the numeric value you want to modify in the Value edit box
- Enter the new label in the Description edit box.
- Press <Enter> or click on the **Add** button.

SIMSTAT will ask you to confirm the replacement of the existing value label. Click on the **Yes** button to replace the existing value label.

USING AN EXISTING VALUE LABELS DEFINITION

Often, several variables share the same value labels. For example, some questionnaires may use a common ordinal scale for several questions. Rather than retyping the same value labels over and over again, SIMSTAT enables the user to establish a link between a variable without value labels and an existing one which already contains labels.

To establish such a link, click on the down arrow button of the **Linked to** list box and select from the list of variables the one containing the value labels you want to use. These labels will appear in the value labels list box.

Once a link has been established, every modification made to value labels list will affect the labels associated with the original variable as well as all other variables currently linked to this variable.



You may also use this feature to copy value labels from one variable to another. To do this, follow the previous instructions to link the current variable to the one from which you want to copy labels. The labels should appear in the value labels list. Then, remove this link by setting the **linked to** option to none.

Entering and Editing Data

Data can be entered and edited directly within SIMSTAT with the use of a data sheet. Here are some quick descriptions of the steps needed to perform specific data editing tasks:

To enter or edit a value:

- Type in the value.
- Press <Tab> to move one cell to the right.
- Press <Enter> to move one cell down.

To erase a value:

- Position the cursor on the cell you want to erase and press the or the <Backspace> key.

To add a row:

- Move to the last row, then press the <Dn> arrow key.

To delete a record:

- Position the cursor on the record (or row) you want to delete.
- Choose DELETE RECORD from the DATA menu.

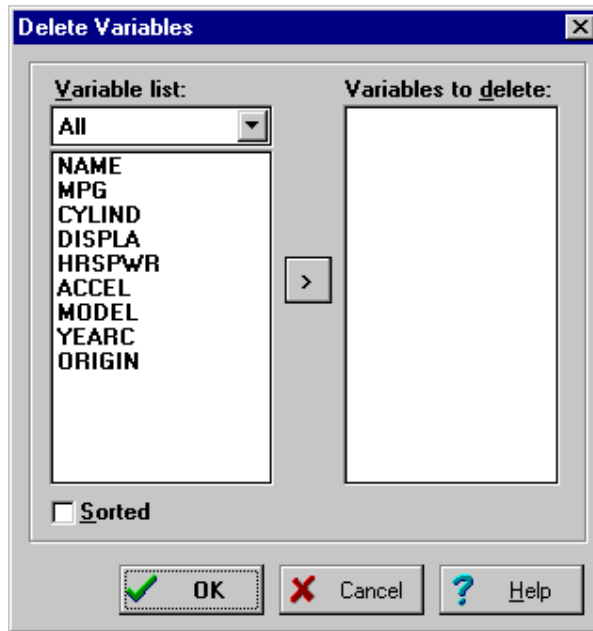
NOTE: You should be aware that the record is not physically removed from the file but is simply tagged as deleted and hidden to the user. To permanently remove the deleted records, use the DATA | PACK FILE command from the FILE menu.



To add new variables:

- Choose the ADD VARIABLES command from the DATA menu. A dialog box will appear that is similar to the one used to create variables for a new data file.
- Specify the name of the new variables and set the variable types, sizes and descriptions as described previously on page 24.
- Click on the **OK** button to create these new variables and add them at the end of the current data set.

To delete existing variables:

- Choose the DELETE VARIABLES command from the DATA menu. The following dialog box will appear.



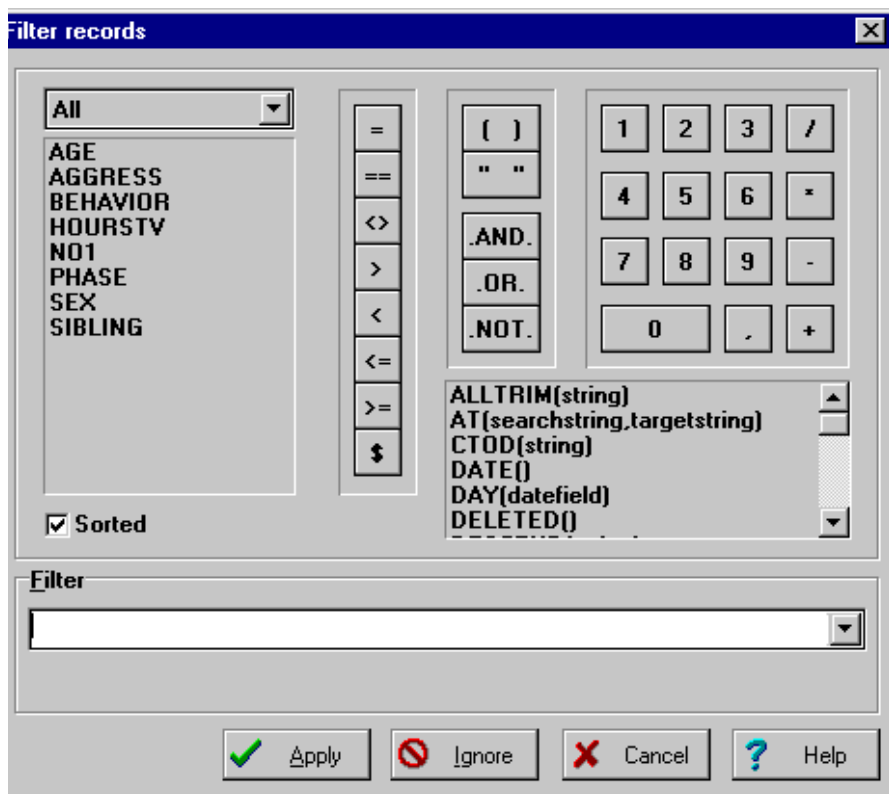
- Highlight the names of the variables you want to delete and click on the  button to move them to the Variables To Delete list box.
- To delete successive variables, click on the first variable, drag the mouse cursor down the list to highlight multiple variables, and then click on the  button.
- Click on the **OK** button to proceed to the deletion of all variables listed in the Variables To Delete list box.

Filtering records

The FILTER RECORDS command temporarily selects cases according to some logical condition. You can use this command to restrict your analysis to a subsample of cases or to temporarily exclude some subjects. You may also use this feature to perform data transformations on a subsample of cases. The filtering condition may consist of a simple expression, or include many expressions related by logical operators (i.e. AND, OR, NOT).

The condition expression should be a valid xBase expression evaluated as true or false and may not exceed 240 characters. To obtain more instructions on expression operators, evaluation rules and supported xBase functions see Appendix A.

When selecting the FILTER RECORDS command, a filter builder dialog box is displayed (see below). You can directly type the filtering expression in the Filter edit box using the proper syntax, or use any elements displayed on the upper part of the Filter Records dialog box to build a valid expression.



To restore previously used filtering conditions, click on the down arrow button located to the right of the Filter edit box.

Once a filtering condition has been entered you can apply the filter and leave this dialog box by clicking on the **APPLY** button. If the filter expression is invalid, a message is displayed and the exit is not performed.

To temporarily deactivate the current filter expression, click on the **IGNORE** button. The filter string will be kept in memory and may be reactivated by choosing the **FILTER RECORDS** command again.

To exit from the dialog box and restore the previous active filtering condition, click on the **CANCEL** button.

Building a valid filtering expression

The upper part of the Filter dialog box contains various elements to help you build a valid filtering expression:

Variable name list box - Double-clicking a variable name from the list box located to the left of the dialog box inserts that name in the edit box at the current caret position.

Function list box - A list of valid xBase expression is displayed to the right of the dialog box. Double-clicking on a xBase function in the list box, inserts that function at the current caret position. When a function requires one or several arguments, the argument section remains highlighted. To replace the highlighted text with a value, an expression or a variable name, simply type the proper text on the keyboard or select a variable name or function.

Numeric, boolean and relational operators buttons - Clicking on any relational or boolean operation or on any numeric button inserts the corresponding symbol in the edit box at the current caret position.

Usually, when part of the filtering expression is highlighted, pressing any keyboard key, or clicking on any variable name or numeric button will replace the highlighted text with the character or expression associated with this key or button. However, when choosing a function requiring a parameter enclosed between parentheses, the highlighted text will not be deleted but will be used instead as the new parameter of this function.

Examples of a simple filtering expression

Here is an example of a simple filtering condition:

```
GROUP = 1
```

This expression tells SIMSTAT to select only those records which have a value for **GROUP** that is equal to 1.

Using boolean operators in the filtering condition

The `.AND.` and `.OR.` switches are logical operators that allow the evaluation of 2 or more logical expressions.

.AND. - The `.AND.` operator instructs SIMSTAT to include only the cases for which both expressions are true. For example,

```
GROUP = 2 .AND. INCOME > 28,000
```

will cause the program to select only those cases from GROUP number 2 that have an INCOME greater than \$28,000.

.OR. - The `.OR.` operator instructs SIMSTAT to evaluate both logical expressions and to include cases for which either expression is true. For example,

```
GROUP = 2 .OR. INCOME > 28,000
```

will cause the program to select cases from GROUP 2 and cases from other groups only if they have an INCOME greater than \$28,000.

.NOT. - The `.NOT.` boolean operator can be used to negate a condition or exclude records meeting a specified criteria. For example,

```
.NOT. GROUP = 1
```

will cause SIMSTAT to exclude all records for which GROUP equal 1.

Multiple level expressions

Multiple parentheses levels can be used to control the order in which expressions are evaluated, as in the following example:

```
(GROUP = 2 .AND. INCOME > 28,000) .OR. (SEX = 1 .AND. EDUC = 5)
```

Sorting records

SIMSTAT provides two different ways to arrange the records (or cases) of a data file in numeric or alphabetic order.

Simple sort

The easiest and quickest way to sort the records on the values of a single variable is to double-click on the variable's name displayed on the top row of the grid. The first time you double-click on a variable's name, the records are immediately sorted in ascending order on this variable. Double-clicking a second time on the same column title sorts the records in descending order.

Complex sort

It is also possible to sort the records of a file using several variables. To achieve this, select the SORT RECORDS command from the DATA menu. When activated, this command brings a dialog box similar to the one used for filtering records. This dialog box allows you to create a custom sorting expression by selecting information and elements displayed on the form. This expression can include almost any supported xBase function.

The sort builder dialog box contains the following elements:

Variable name list box - Double-clicking on a variable name from the list box located on the left of the dialog box inserts that name in the edit box at the current caret position.

Function list box - A list of valid xBase expressions is displayed to the right of the dialog box. Double-clicking on an xBase function from the box inserts that function at the current caret position. When a function requires one or more arguments, the argument section remains highlighted. To replace the highlighted text with a value, an expression or a variable name, simply type the proper text on the keyboard, or select a variable name or function.

Relational operations and numeric buttons - Selecting any relational operation or numeric button inserts the corresponding symbol in the edit box.

Usually, when part of the sorting expression is highlighted, pressing any keyboard key or clicking on any variable name or numeric button will replace the highlighted text with the character or expression associated with this key or button. However, when choosing a function requiring a parameter enclosed between parentheses, the highlighted text will not be deleted but will be used instead as the new parameter of this function.

Once a valid sorting expression has been entered, you may use the Sort Order option to specify whether the records should be sorted in ascending or descending order.

To execute the sorting and leave the dialog box, click on the **APPLY** button. If the sorting expression is invalid, a message is displayed and the dialog box remains displayed.

To temporarily deactivate the current sorting expression, click on the **IGNORE** button. The sorting string is kept in memory and may be reactivated by choosing the **SORTING RECORDS** command again.

Press the **CANCEL** button to abort the current sorting operation, leave the dialog box and restore the previous sorting order.

Examples:

When the following sorting expression is used,

```
BIRTHDAY
```

and the Ascending radio button is activated, the records are sorted in ascending order by the variable birthday.

In the following sorting expression:

```
SEX*100+DESCEND(AGE)
```

where the variable **SEX** is a numeric variable (0 = male and 1 = female), **SIMSTAT** sorts the records with male subjects at the top of the data grid, ordered from the oldest to the youngest, and all females at the bottom of the grid in the same order.

Importing data from other applications

SIMSTAT allows you to import data files from other statistical programs, spreadsheet and database applications, as well as from plain ASCII data files (comma or tab delimited text). The program can read data stored in the following formats:

SIMSTAT for DOS (v1.0 - v3.5)
SPSS/PC+
SPSS for Windows
PARADOX (v3.0 - v5.0)
LOTUS 123 (v1.0 - v5.0)
SYMPHONY (v1.0 - v1.1)
EXCEL (v2.1 - v5.0)
QUATTRO PRO (v1.0 - v6.0)
COMMA SEPARATED VALUES (Windows and DOS)
TAB SEPARATED VALUES (Windows and DOS)

To import data from any of these applications:

- Choose the **DATA | IMPORT** command from the **FILE** menu.
- Select the file format using the List File of Type drop down list.
- Select the file you want to import and click **OK**.

You will find below detailed information about specific file formats such as their formatting requirements (if any), the supported features and limitations.

SIMSTAT for DOS files

SIMSTAT imports files from its DOS version with .SIM extensions. The program will import variable and value labels as well as user missing values associated with each variable.

SPSS/PC+ and SPSS for WINDOWS files

SIMSTAT reads compressed and uncompressed SPSS/PC+ system files with .SYS extensions. The program also supports variable and value labels as well as user missing values associated with each variable. SPSS/PC+ files are saved in uncompressed format.

SIMSTAT can also read compressed and uncompressed SPSS for Windows data files (.SAV extension) containing less than 1022 variables. The program will import variable and value labels as well as user missing values associated with each variable.

ASCII data files

SIMSTAT will read up to 500 numeric and alphanumeric variables from a plain ASCII file (text file). The file must have the following format:

- Every line must end with a carriage-return.
- The first line must include the variable names, separated by spaces, tabs and/or commas.
- Variable names may have a length of not more than 10 characters. Longer strings are truncated at 10 characters.
- The remaining lines must include the numeric scores separated by spaces, tabs and/or commas.
- Each line must contain scores for one case and variables must be in the same order for all cases.
- All invalid scores and all blanks encountered between commas or tabs are treated as missing values. A single dot can also be used to represent a missing value.
- Comments can be inserted anywhere in the file by putting an asterisk ('*') at the beginning of the line.
- Blank lines can also be inserted anywhere in the file.

Spreadsheet data files

SIMSTAT reads spreadsheet files produced by LOTUS 1-2-3 (v1.1 to v5.0), SYMPHONY (v1.0 and v1.1), EXCEL (v2.2 to v5.0), and QUATTRO PRO (v1.0 to v6.0). When a spreadsheet data file is selected, the program displays a dialog box where you can specify, in the case of multiple page spreadsheets, the page where data are located, and the range of cells to be read. You must specify a valid range name or provide upper-left and lower-right cells, separated by two periods (such as A1..H20). If you set the Range Name list box to ALL, the program attempts to read the whole file.

The selected range must be formatted such that the columns of the spreadsheet represent variables while the rows represent cases or records. Also, the first row should preferably contain the variable names while the remaining rows hold the data, one case per row.

Cells in the first row of the selected range are treated as field names. SIMSTAT will automatically determine the most appropriate format based on the data it finds in the worksheet columns. If no variable name is encountered, SIMSTAT will automatically provide one for each column in the defined range.

When reading the data for analysis, all blank cells or cells that do not correspond to the variable type (e.g., alphanumeric entries under a numeric variable, or a numeric value under a string variable) are treated as missing values.

Exporting data to other applications

SIMSTAT allows you to export data to other applications, including:

SIMSTAT for DOS
SPSS/PC+
DBASE
PARADOX (v3.0 - v5.0)
LOTUS 123 (v1.0 - v5.0)
SYMPHONY (v1.0 - v1.1)
EXCEL (v2.1 - v5.0)
QUATTRO PRO (v1.0 - v6.0)
COMMA SEPARATED VALUES (Windows and DOS)
TAB SEPARATED VALUES (Windows and DOS)

When exporting a data file, SIMSTAT will use the current filtering condition to determine which records will be exported. The active sorting order will also be used to control the record sequence in the new file format.

To export data to any of these applications:

- Set the filtering and sorting conditions of the active data file to display the records as they should be exported.
- Choose the DATA | EXPORT command from the FILE menu.
- Select the file format you want to create using the Save As File Type drop down list.
- Type a valid filename with the proper file extension.
- If necessary, use the Directories and Drives boxes to specify the storage location of your choice.
- Click on the **OK** button.

You will find below a list of available file formats along with information about the export function limitations.

SIMSTAT for DOS (v1.0 - v3.5) & SPSS/PC+

- A maximum of 500 variables can be exported.
- Only the first user missing value is kept.
- If an alphanumeric field is longer than 8 characters, then all values will be truncated to 8 characters.

ASCII data files

- A maximum of 500 variables will be exported.
- Variable descriptions, missing values definition, and value labels are not supported.

All spreadsheet formats

- A maximum of 255 variables will be exported.
- When exporting to a file format supporting multiple pages, the data are saved to the first page.
- Variable descriptions, missing values definition, and value labels are not supported.

Creating a file with subsets of cases

The export feature can also be used to create a new data file with an identical structure as the current data file but with only a subset of cases.

To create such a file:

- Set the filtering and sorting conditions of the active data file to display the records as they should be saved in the new data file.
- Choose the DATA | EXPORT command from the FILE menu.
- Set the Save As File Type drop down list to DBASE
- Type a valid filename including the .DBF file extension.
- If necessary, use the Directories and Drives boxes to specify the storage location of your choice.
- Click on the **OK** button.

Merging data files

There are circumstances when two or more data files need to be merged together. For example, you may need to append data from additional subjects or from another time period to the end of an existing data file or you may need to add new information on existing individuals by adding new variables. Simstat provides two merging methods to either add new records or new variables to an existing data file.

To add new records to an existing data file

- Open the data file to which you would like to append records.
- Choose the DATA | APPEND | RECORDS command from the FILE menu.
- Select the data file to be merged with the current file, and click **OK**.

All variables with matching names and types are appended to the current data file. If variable length differs, data in the merged database is either truncated or padded with spaces.

Deleted records in the merged file are not appended to the current database.

To add new variables to an existing data file

- Open the data file to which you would like to append new variables.
- Choose the DATA | APPEND | VARIABLES command from the FILE menu.
- Select the data file to be merged with the current file, and click **OK**.
- Choose among the list of variables shared by both data files, those that should be used to perform the record matching and click **OK**.

If no record of the secondary file matches the key values in the current file, missing values are assigned to the new variables. If several records in the secondary file match the key values, values are extracted from the first matching record.

Variables with identical names are ignored. To replace the values of one of those variables with the ones in the second data file, simply delete this variable from the current data file prior to the APPEND | VARIABLES operation.

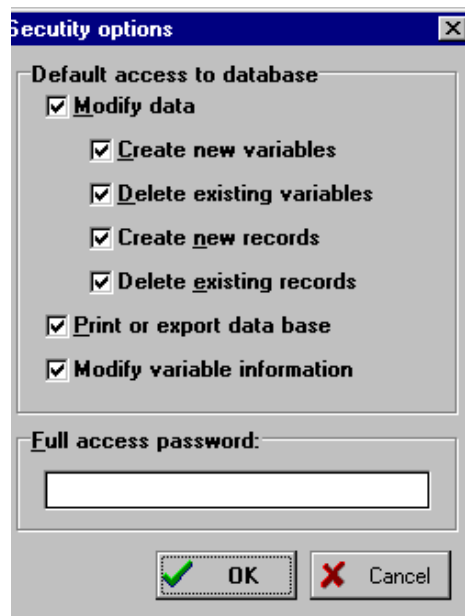
Limiting access to data files

Access privileges are useful for controlling access to sensitive data in a file, whether or not the file is shared on a network. For example, you may want to prevent modification of data files by other users, or simply prevent accidental modification by yourself.

SIMSTAT allows restrictions to be placed on the type of activities that can be performed on a data file. Full access to the data file can be regained by providing a single password.

To define the default limited access to a data file

- Open the file you want to restrict access to.
- Choose the DATA | SECURITY command from the FILE menu. The following Security dialog box will appear:



When first accessed, all options are checked, indicating that no restriction to the data file has been defined.

- Disable the privileges you want to associate with the password. For example, if you disable the Delete Existing Variables and Delete Existing Records options, it will be possible for any user to edit data, create new variables, or modify the variable definitions (missing values, labels, display widths, etc.). However, they will not be able to delete any existing record or variable in the data file.

- Type the password that will give full access to the data file and click on the **OK** button to exit the dialog box.

NOTE: You can limit default access for users without the need to provide a password by leaving the Full Access Password edit box empty. By doing this, the access will be limited every time the data file is opened, but it will still be possible for anyone to modify this access privilege by altering the options in this Security dialog box.

To gain full access to a data file with restricted access.

- Choose the DATA | SECURITY command from the FILE menu and enter the proper password to display the Security dialog box.
- Enable the options you need and click on the **OK** button to exit the dialog box.

CAUTION: All options modified when regaining full access become the new default access values. You should take care to reset these values before closing the data file if you want to maintain this default limited access to the data file.

Archival backup of data files

During a research project, data files are frequently modified (cleaning, recoding, mathematical transformation, etc.). At some point you may need to return to a previous version of an existing data file to recover lost variables or cases that have been transformed or deleted, or just to make routine verifications. In order to do this, several successive backup copies of this data file should be created and kept. Another reason to make backup copies of data files is to prevent the accidental loss of an entire data file caused by a hardware failure or software malfunction.

SIMSTAT provides a simple archiving procedure that allows one to quickly create backup copies. This procedure stores a copy of the currently active data file and all its related files (structure definition, value labels, variable sets, etc.) in a single compressed file. SIMSTAT uses the industry standard Zip format as its own archive file format allowing you manage those archive files outside of SIMSTAT using any application than can manipulate these files. Also, the high level of compression achieved on typical SIMSTAT data files (usually more than 90-95%) allows you to create backup copies of large data files on a single diskette, or keep several copies of your data file on your hard drive without sacrificing precious disk space.

Another benefit of this archiving feature is to facilitate the transfer of your data file and all its related files to another computer by creating a single file that includes all related files, and can easily fit on a single diskette.

To make an archived copy of the current data file

- Select the DATA | ARCHIVE | BACKUP command sequence from the FILE menu. A Save File dialog box will be displayed.
- Set the drive and directory setting to the location where you want to store the compressed file.
- Enter a valid file name and click **OK**

To restore a file from an archived copy

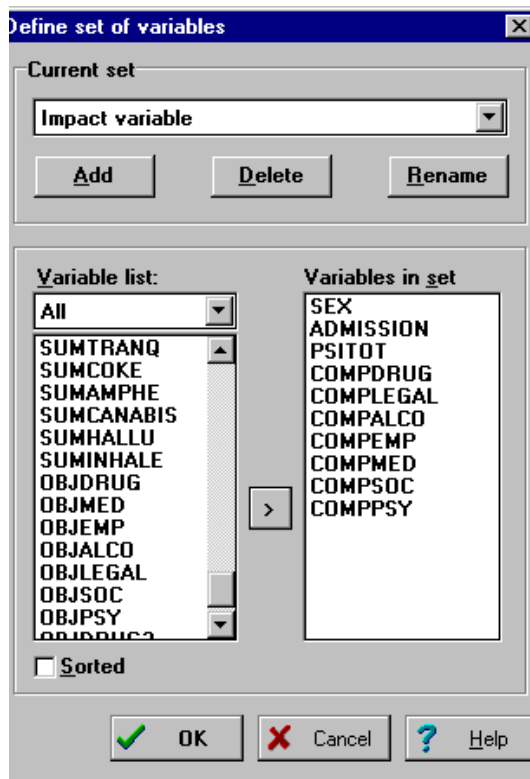
- Select the DATA | ARCHIVE | RESTORE command sequence from the FILE menu. An Open File dialog box will appear displaying all ZIP files.
- Select the proper ZIP file and click on the **OK** button. A second dialog box will appear to let you specify the location where data should be restored. The default location is the same directory as the archive file.
- If needed, change the drive and directory setting and click on the **OK** button to proceed with the extraction.

Creating and using variable sets

When working with large data files involving several hundred variables, it may become difficult to locate the variables of interest. The SETS OF VARIABLES command facilitates this task by allowing you to define several groups of variables in the data file and later restrict the variables displayed in dialog boxes involving selection of variables to those associated with a specific set.

To define a set of variables

- Select the SETS OF VARIABLES command from the DATA menu. The following dialog box will appear.



- Click on the **ADD** button and enter the new set name. You can enter a string of up to 20 characters to describe this new set.
- Move all the variables that will make up the new set to the list box on the right of the screen.

- Click on the **OK** button to exit from the dialog box or click on the **ADD** button again to create another set.

To modify a set definition

- From the Current Set drop down list box, select the set you want to modify.
- Add or remove the variables from the set.

To rename a set of variables

- From the Current Set drop down list box, select the set you want to rename.
- Click the **RENAME** button and enter the new set name.
- Click on the **OK** button to confirm the new name.

To delete a set

- From the Current Set drop down list box, select the set you want to delete.
- Click on the **DELETE** button and enter the new set name.

4 - DATA TRANSFORMATION

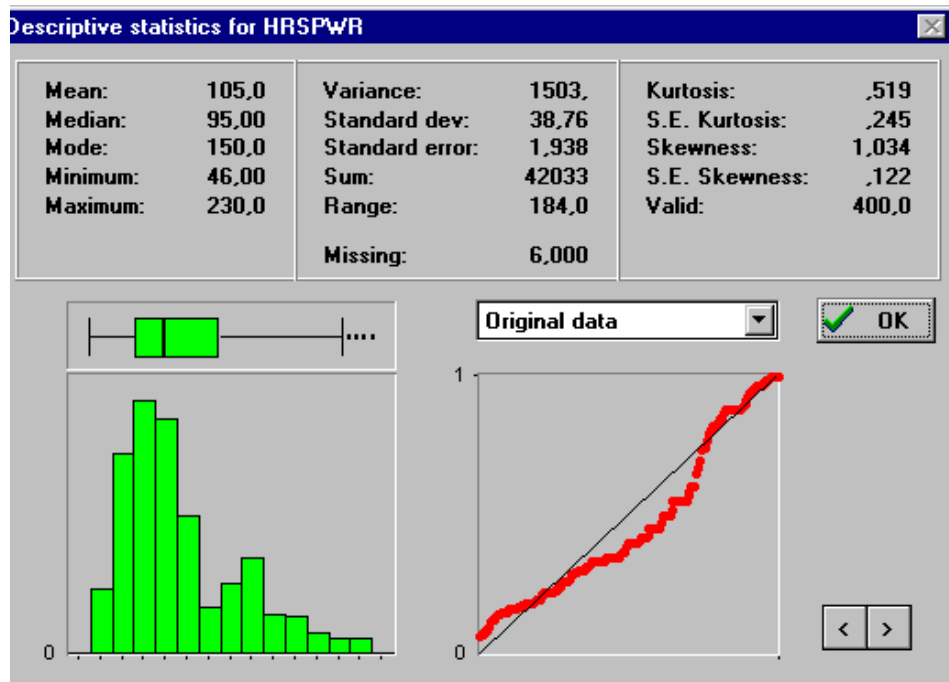
Often you may need to create new variables to represent global scores based on several responses given by a subject or to create composite indices based on several indicators. Preliminary analysis of existing variables may also reveal either coding errors, or an inadequate data distribution for the kind of analysis you want to perform. SIMSTAT provides several powerful commands to transform values of existing variables or create new variables by performing mathematical operations on existing ones. This chapter presents data transformation features currently available in SIMSTAT.

The Transformation submenu contains various commands to perform transformation of existing variables or to create new variables from computation on existing ones.

Compute	Provides transformations using various functions on one or more variables.
Quick transform	Performs immediate transformation of the current variable using various functions.
Recode	Changes the values of numeric variables, or collapses values of a continuous variable into categories.
Rank	Replaces the values of the current variable with their rank.
Dummy coding	Automatically creates as many dummy variables as needed to represent the values of a nominal variable.
Numeric coding	Automatically creates a numeric variable to express the values of a string variable.

Examining data distributions

SIMSTAT provides a special feature that allows you to inspect the distribution of a variable and assess the effect of common transformations on the data distribution. To use this feature, position the data grid cursor on the numeric or date variable you want to examine and select the VARIABLE STATISTICS command from the DATA menu. A dialog box will appear displaying various statistics including mean, median, standard error of the mean, variance, standard deviation, skewness, kurtosis, minimum and maximum values, etc., as well as 3 graphs (i.e. a box-plot, a histogram and a normal probability plot).



By default, the distribution statistics and the graphs displayed depict the untransformed values of the variable. A list box allows one to temporarily transform those values using common transformation formula and examine the resulting distribution.

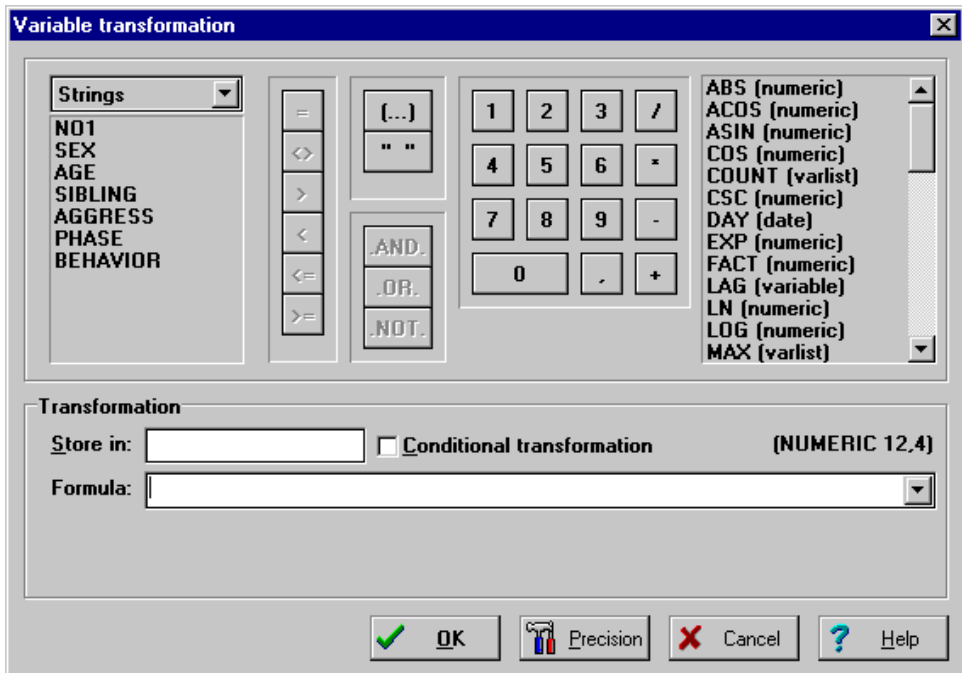
Navigation keys at the lower right side of the window allow you to quickly move from one variable to another.

To perform a transformation on your data, see section COMPUTING VALUES or QUICK TRANSFORM below.

Computing values

The COMPUTE command allows the transformation of existing variables or the creation of numeric variables by performing numeric transformations on existing variables. This command offers more than 50 operations and functions including numerical operators, trigonometric transformations (cos, sin, log, etc.), statistical functions (mean, minimum, maximum across variables or cases, etc.), date, and random number operations. Conditional transformations can also be performed using an IF-THEN-ELSE logical structure.

To perform a numeric transformation, select the TRANSFORM VARIABLE | COMPUTE command from the DATA menu. When this command is evoked the following dialog box appears:



Store in - This edit box allows you to specify the target variable where the computation results will be stored. If the target variable already exists, the program will ask you if you want to replace its values with those produced by the transformation. If the variable does not exist, the program will ask you to confirm the creation of this new variable. If you answer yes to this question, the new variable will be appended to the left end of the data sheet. (see “Setting Program Preferences” on page 235 for information on how to disable these confirmation dialogs). When a new numeric variable is created, the program needs to know in advance its physical size and precision. The current default size and number of decimal places used to store the result of a transformation is

indicated in the upper right corner of the Transformation group box. To modify these values, click on the precision dialog box and enter the new default size and number of decimal places.

Conditional transformation - This check box allows you to choose between a simple computing formula or a conditional transformation. When this option is disabled, a single edit box (i.e., Formula) is shown and the transformation is performed on all currently selected records. Enabling this option allows you to perform a conditional transformation using an IF-THEN-ELSE logical structure. The single edit box is replaced by 3 edit boxes allowing you to specify a logical condition and two transformation formulas. (see “conditional transformation” below).

Formula - This field contains the computing formula used to compute the value of the target variable. This formula can contain existing variable names, numeric constants, arithmetic operators or any supported numeric, trigonometric or statistical functions. You can type the numeric transformation in the Formula edit box directly, using the proper syntax, or use any element displayed on the upper part of the Variable Transformation dialog box to build a valid expression. To restore previously used numeric transformations, click on the down arrow button located at the right of the Formula edit box.

Once a valid target variable and a numeric expression have been entered you can leave this dialog box and perform the transformation by clicking on the **OK** button. If the transformation expression is invalid, a message is displayed and the dialog box remains displayed.

To exit from the dialog box without performing a transformation, click on the **CANCEL** button.

Building a valid transformation expression

The upper part of the dialog box contains various elements to help you build a valid numeric expression:

Variable name list box - Double-clicking a variable name from the list box located to the left of the dialog box inserts that name in the edit box at the current caret position.

Function list box - A list of valid numeric functions is displayed to the right of the dialog box. Double-clicking on a function name in the list box inserts that function at the current caret position. When a function requires one or several arguments, the argument section remains highlighted. To replace the highlighted text with a value, an expression or a variable name, simply type in the proper text or select a variable name or function.

Numeric, arithmetic, boolean and relational operators buttons - Clicking on any relational or boolean operation or on any numeric button inserts the corresponding symbol in the edit box.

Usually, when part of an expression is highlighted, pressing any keyboard key, or clicking on any variable name or numeric button will replace the highlighted text with the character or expression associated with this key or button. However, when choosing a function requiring a parameter enclosed between parentheses, the highlighted text will not be deleted but will instead be used as the new parameter of this function.

You will find below a list of valid symbols and functions along with a short description of each, as well as some additional syntax information.

Arithmetic Operators

You can use any of the following symbols in a transformation formula.

+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponentiation

When a transformation expression is evaluated, exponentiation is performed first, followed by division and multiplication and finally addition and subtraction. Multiple parentheses levels can be used to control the order in which expressions are evaluated as in the following examples:

$2 * 3 + 4 * 2$	returns 48
$2 * (3 + 4) * 2$	returns 28
$((2 * 3) + 4) * 2$	returns 20

Constants

PI	3.1415926535897932385
-----------	-----------------------

Numeric and Trigonometric Functions

Syntax FUNCTION (value, variable or expression)

ABS()	Absolute
ACOS()	Arccosine
ASIN()	Arcsine
ATAN()	Arctangent
CSC()	Cosecant
COS()	Cosine
EXP()	Exponential
FACT()	Factorial
LN()	Natural logarithm
LOG()	Base-10 logarithm
MOD10()	Modulus
RND()	Round
SEC()	Secant
SQRT()	Square root
SQR()	Square
SIN()	Sine

TAN () Tangent
TRUNC () Truncate

Statistical Functions (across variables)

The following statistical functions are computed on several variables for each record individually. For example, if you enter:

```
MEAN(Q1 Q2 Q3 Q4 Q5)
or MEAN(Q1..Q5)
```

SIMSTAT will compute the mean of values stored in the five specified variables of the current record (missing values are excluded). To obtain a statistic on all selected records see the next section “Statistical Functions (across records)”.

Syntax FUNCTION (Var [Var Var..Var])

MEAN () Mean
COUNT () Count (missing value excluded)
SUM () Sum
STDEV () Standard deviation
VAR () Variance
MIN () Minimum
MAX () Maximum
WSUM () Weighted sum (adjusted for missing value)

Statistical Functions (across records)

The following statistical functions are computed on a single variable for all currently selected records. For example, if you enter:

```
VMEAN(AGE)
```

SIMSTAT will compute the mean of the values stored in the AGE variable for all selected records (missing values are excluded). To compute a statistic on several variables for each record individually, see the previous section “Statistical Functions (across variables)”.

Syntax FUNCTION (Variable)

VMEAN () Mean
VCOUNT () Count
VSUM () Sum
VSTDEV () Standard deviation
VVAR () Variance

VMIN()	Minimum
VMAX()	Maximum
ZSCORE()	Normalized score
LAG()	Lag

Random Number Functions

Syntax FUNCTION (value, variable or expression)

NORMAL(X)	Normal pseudo-random number with mean of 0 and standard deviation of X
UNIFORM (X)	Uniform pseudo-random number between 0 and X

Date Functions

Most date functions can be computed on either date or numeric variables. Numeric variables are automatically transformed into a date corresponding to the number of days since 1/1/0001. The TODAY function does not have any argument, while the YRMODA function takes 3 arguments separated by commas. Each arguments can be either a numeric variable, a constant, or an expression. All other functions require a single argument that can consist of a date, a numeric variable, a constant or an expression.

YRMODA()	Argument: (yy,mm,dd). Converts 3 values, variables or expressions into a numeric value expressing the number of days since 1/1/1900.
YEAR()	Returns the year of a date.
MONTH()	Returns the month of a date.
DAY()	Returns the day of a date.
WEEKDAY()	Returns a numeric value representing the day of the week (from 1 to 7)
TODAY()	Returns the current date

Missing Value Constant

When the target variable already exists and a transformation expression is left blank, the value of the target variable remains unchanged. To erase the previous value and leave the variable blank, use the SYSMIS keyword.

SYSMIS	System missing value
---------------	----------------------

Performing conditional transformations

Conditional transformations allow you to apply a numeric transformation to a subset of records meeting a specific criteria. When the **CONDITIONAL TRANSFORMATION** check box is enabled, the single **FORMULA** edit box is replaced with the following 3 edit box:

IF - The **IF** edit box contains a logical expression up to 250 characters long specifying criteria for records on which to apply a specific transformation. This expression must be evaluated as true or false. It can be a simple logical expression including the following 3 components: a variable name, a relational operator, and a variable name or a numeric constant. It can also be a complex expression composed of many simple expressions related with logical operators (**AND**, **OR**, **NOT**). Parentheses can be used to specify the order in which expressions are evaluated. For example:

```
((GROUP = 1) .AND. (AGE > 30)) .OR. (GROUP = 2)
```

This expression can also include any xBase function, provided that the final result of the expressions can be evaluated as true or false (For more information on valid xBase expressions, see appendix A).

Then / Else - The **THEN** and **ELSE** edit boxes contain the transformation expressions you want the program to apply. See the previous section “Building a valid transformation expression” for information on syntax rules and available functions. If the logical expression entered in the **IF** edit box is evaluated as true, the transformation expression in the **THEN** edit box is computed; if not, the expression in the **ELSE** field is used. If an edit box contains no transformation expression, the existing value remains unchanged. For example, the following setting:

```
STORE IN: HEIGHT
      IF: SEX = 'Female'
      THEN: HEIGHT * 1.3
      ELSE:
```

Will replace, for each record where the character variable **SEX** is equal to “Female”, the current value stored in variable **HEIGHT** with this same value multiplied by 1.3; all other records remain unchanged.

Quick transformations

The QUICK TRANSFORM command allows immediate transformation of the current variable using various functions.

Z Score	$ZSCORE(X)$ or $(X - VMEAN(X)) / VSTDEV(X)$
Remove mean	$X - VMEAN(X)$
Reverse	$VMAX(X) - X + 1$
Add constant...	$X + \text{constant}$
Square root	$SQRT(X)$
Natural logarithm	$LN(X)$
Inverse	$1/X$

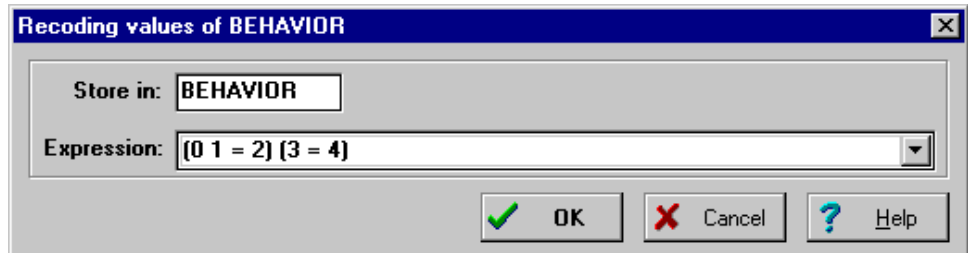
For the square root, natural logarithm and inverse transformations, if the variable contains values less than 1, SIMSTAT asks whether to add a constant in order to bring this minimum to 1.

Recoding values of a variable

The RECODE command provides an easy way to make multiple changes to the values of numeric variables, or to collapse values of a continuous variable into categories. To activate this command:

- Position the data sheet cursor on the variable you want to recode
- Select the TRANSFORM VARIABLE | RECODE command from the FILE menu.

The following dialog box will appear:



Store in - This edit box allows you to specify the target variable where the computation results will be stored. By default, the value is set to the name of the variable containing the values that will be used for recoding. To store the result in another variable, change this text box to the new target variable name. If you keep the original variable name or specify an existing variable name, the program will ask you if you want to replace its values with those produced by the transformation. If the variable does not exist, the program will ask you to confirm the creation of this new variable.

Expression - The recode expression can consist of several transformations enclosed in parentheses, each including one or more values (or a value range), an equal sign, and the new value. For example:

```
(1 = 2) (2 = 1) (3 4 5 = 3) (6..10 = 4) (ELSE = SYSMIS)
```

Each value on the left of the equal sign is recoded into the value on the right. The recoding proceeds from left to right and stops after a transformation occurs. The SYSMIS and MISSING keywords can be used to represent missing values, while the ELSE keyword represents all non specified values.

Once a valid target variable and a recoding numeric expression have been entered you can leave this dialog box and perform the recoding by clicking on the **OK** button. If the recoding expression is invalid, a message is displayed and the exit is not performed.

To exit from the dialog box without performing the recoding, click on the **CANCEL** button.

Transforming a variable into ranks

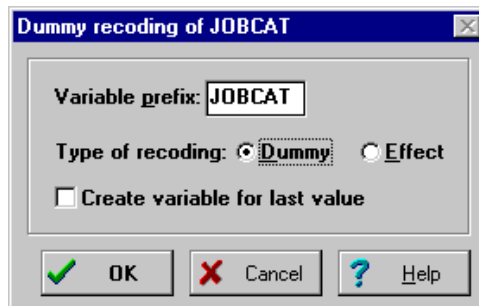
The RANK command replaces the values of a variable by their rank. If ties occur, the mean rank of tied values is used. Missing values are excluded.

To transform a variable into ranks:

- Position the data sheet cursor on the variable you wish to transform into ranks.
- Select the TRANSFORM VARIABLE | RANK command from the DATA menu.
- Enter the new variable name under which the rank should be stored and click **OK** to proceed to the transformation.

Dummy recoding of nominal variables

The DUMMY RECODING command automatically creates as many dummy variables as needed to represent the values of a nominal variable. To create such dummy variables, position the cursor on the nominal variable you wish to recode and select the TRANSFORM VARIABLE | DUMMY RECODING command from the DATA menu. The following dialog box will appear:



Variable prefix - The variable prefix option allows you to enter a prefix that will be used to generate dummy variable names. The maximum length of this prefix is 7 characters. Numeric values going from 1 up to the number of variables required will be appended to this prefix.

Type of recoding - This option provides a choice between two different methods for recoding nominal variables.

Dummy - The dummy (or binary) coding involves the creation of dichotomous vectors to represent membership in the categories of nominal variables such that subjects in a given category are assigned 1, while non members are given a score of 0. The number of dummy variables needed to contain all the information of a nominal variable equals the number of different values in this variable minus one. While most analyses involving dummy variables require you to enter only those $g-1$ dichotomous variables,

you may need to perform separate analysis for each category. When the dummy coding type is chosen, a check box appears allowing you to create this last dichotomous variable. For example, if you recode a variable containing 3 different values, setting this check box will create 3 dummy variables, one for each group. If this option is disabled, the procedure will create only 2 variables.

Effect coding - Effect coding is similar to dummy coding, except that the last group is assigned -1s in all vectors instead of 0s.

When you click on the **OK** button, the program reads all active records and asks you to confirm the computation of the required number of dichotomous variables. If variables with similar names exist in the database, the confirmation dialog box will display the number of existing variables that would be overwritten as well as the number of variables that need to be created by this command. Click on the **YES** button to confirm the creation and overwriting of variables. Click on **NO** to abort the procedure.

To exit from the dialog box without performing the recoding, click on the **CANCEL** button.

Numeric recoding

Most statistical analyses require variables that can be expressed as numeric values (such as numeric, date, and logical data type). The NUMERIC CODING command automatically creates a numeric variable to express the values of a string variable. To create this numeric variable, position the cursor on the string variable and select the TRANSFORM VARIABLE | NUMERIC RECODING command sequence from the DATA menu.

When this command is evoked, a dialog box appears asking you the name of the new numeric variable. If you specify a non existing variable, the program will ask you to confirm the creation of this new numeric variable. If you click on the **YES** button, this variable will be created and will contain integer values representing each unique alphanumeric value found in the original string variable. Value labels corresponding to each numeric value will be automatically created and stored with the variable.

If the name of an existing variable is provided, you will need to confirm the overwriting of values in this variable. When the variable name supplied is the same as the original string variable, the program assumes that you want to transform the current string variable into a numeric variable and will erase the original variable and add a new numeric variable of the same name. This new variable will be located at the left end of the data worksheet.

5 - WORKING WITH THE NOTEBOOK

The Notebook window displays the statistical output for all analyses performed during a session. The notebook metaphor provides an efficient way to browse and manage outputs. The text output of each analysis is displayed on a separate page. You can turn pages with the mouse by clicking on the page corner icons at the bottom of the notebook, or by using the <PgUp> and <PgDn> keyboard keys. Tabs may also be added to create sections in the notebook, allowing you to store different types of analyses in different sections of the notebook. An index of all analyses included in the notebook is also provided. This index can be used to quickly locate and go to a specific page, move pages within the notebook, or delete some pages. While each page can be annotated or edited, it is also possible to add empty pages for recording ideas or remarks, sketching an analysis plan, or writing down interpretation of results.

Notebook: UNTITLED Page 3

FREQUENCIES: CYLIND

Value	Frequency	Percent	Valid Percent
3	4	1.0	1.0
4	207	51.0	51.0
5	3	.7	.7
6	84	20.7	20.7
8	108	26.6	26.6
TOTAL	406	100.0	100.0

Mean	5.475	Variance	2.931	Kurtosis
Median	4.000	Std Dev	1.712	S.E. Kurt
Mode	4.000	Std Err	.085	Skewness
Minimum	3.000	Sum	2223.000	S.E. Skew
Maximum	8.000	Range	5.000	Valid

Analysis Index

The following section provides basic instruction to:

- navigate in the notebook and edit its contents.
- use tabs and the index to manage output.
- add or remove notebook pages.
- round numeric values.


Navigating in the notebook


To move the caret to a specific location, you can click with the mouse on that location. You can also use the following keys to navigate in the notebook:

<Up>	Move one line up.
<Down>	Move one line down.
<PgUp>	Move up one screen. If the cursor is on the last line of the page, pressing this key moves to the next page.
<PgDn>	Move down one screen. If the cursor is on the first line of the page, pressing this key moves to the previous page.
<Home>	Move to the beginning of the line.
<End>	Move to the end of the line.
<Ctrl-Home>	Move to the first line.
<Ctrl-End>	Move to the last line.
<Ctrl-Right>	Move to the beginning of the next word.
<Ctrl-Left>	Move to the beginning of the previous word.
<Ctrl-Enter>	Insert a page break

It is possible to locate a specific string on the current page or in the entire notebook by choosing the FIND command in the EDIT menu or by pressing **<Ctrl-F>**.


You can also use the page flip icons located at the bottom of the notebook to move within the notebook pages.

Left click on the  icon to move to the next page or click on the right button to move to the previous page.

Left click on the  icon to move to the previous page. Pressing with the right button moves you to the previous page.

To add an empty page

Empty pages can be added in the notebook. Those pages may be used to sketch an analysis plan, write down some comments or interpretations of analysis results, or simply put some reminders of things to do. To add an empty page to your notebook:

- Make the notebook the active window.
- Select the page after which you want the empty page to appear.
- Select the PAGES | NEW command from the EDIT menu or click on the  button.

If the current page is the first page of a section, a dialog box will appear giving you a choice to enter the new page before or after the current page.

To erase existing pages

- Select the PAGES | DELETE command from the EDIT menu.
- Specify the range of pages you want to delete and select **OK** to proceed.

Note: You may also use the notebook index to delete single pages or to erase an entire section.

Using the notebook index

SIMSTAT automatically creates an index of all analysis output stored in the notebook. This index is useful to get an overall view of the notebook content and to quickly locate and move to a specific page. It also allows you to restructure the contents of the notebook. To view the notebook index, click on the Index tab at the bottom of the notebook.

Expanding and collapsing sections

- Choose a section in the outline.
- Press <Enter> or double click on the section name or its folder icon.


To locate and display a specific page

- Browse through the index until you find the page you want to display.
- Double-click on the page name or icon.

To move a page within the notebook

- Position the mouse cursor over the page you want to move.
- Press and hold down the mouse button.
- Drag the page over the page just before its new location and release the mouse button.


To erase a page

- Select the page you want to erase.
- Select the DELETE command from the EDIT menu or click on the  button.

Output management using Tabs

A typical research project or analysis task can result in hundreds, or even thousands of pages of statistical output. Using tabs can facilitate the management of your output by allowing the creation of sections in the notebook. For example, you may choose to create sections for different types of analyses (descriptive analysis of data, reliability analysis of instruments, regression analysis to test your hypothesis, etc.), or for analysis performed on different subsamples or at different time periods, etc.

To create a new tab


- Make sure the Notebook window is the active window.
- Select the TABS | ADD command from the EDIT menu or click on the  icon.
- Enter the new tab name and its page location.
- Press <Enter> or click on the **OK** button to create the new tab.

Once your tabs are created, you can quickly move to a specific section by clicking on its tab at the bottom of the notebook. When activated, a section become the default output location. This means that the output of all analyses performed while this section is active will be appended to the end of this section.

To delete an existing tab (but not its contents)

- Click on the tab you want to delete or select any page in its section.
- Select the TABS | DELETE command from the EDIT menu.

To delete an entire section

- Click on the Index section to display the notebook index.
- Select the tab of the section you want to delete.
- Select the ERASE command from the EDIT menu or click on the  icon.

To modify a tab name or change its location

- Click on the tab you want to modify or select any page in its section.
- Select the TABS | EDIT command from the EDIT menu.
- Modify the tab name or its page number and click **OK** to confirm the changes.

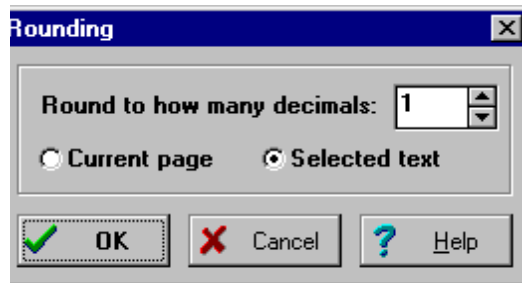
Suggestion

It is advisable to keep a copy of the data used with each analysis output. To achieve this, you can create a special DATA section and use the LIST command to create listings of all data sets used in your analysis. It may also be a good idea to keep a log of the analysis commands and options used with your statistical results. To achieve this, simply use the RECORD SCRIPT feature to automatically generate those commands and copy the content of the script window to an empty notebook page.

Rounding numerical values

SIMSTAT provides a feature that allows you to reduce the number of decimal places of numeric values within a selected area or in the active page of the Notebook. To round the values, perform the following steps:

- If necessary, select the portion of text containing the numeric values you want to round.
- Select the ROUND command from the EDIT menu. The following dialog box will appear:




- Specify the number of decimal places to which you want to round numbers.
- Set the radio button to Current Page or Selected Text depending on whether you want to round values for the entire page or only in the selected portion of it.
- Select the **OK** button to exit the dialog box.

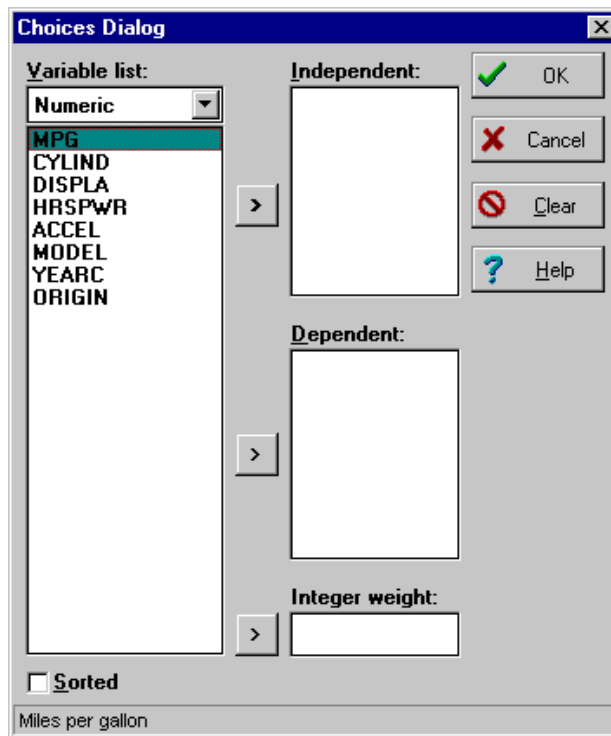
6 - STATISTICAL ANALYSIS

This reference chapter provides a description of the Variable Selection dialog box followed by a description of every statistical analysis that appears in the STATISTICS menu. To allow you to quickly find the information you need, the statistical commands are presented in alphabetical order. The sole exceptions to this rule are the BOOTSTRAP resampling procedures that are described under the BOOTSTRAP heading.

Assigning variables for statistical analysis


Use CHOOSE X-Y from the STATISTICS menu to select the variables to be used in subsequent analyses. You can also press the <F3> function key or click on the  button to access this function.

When this command is selected, the program displays a dialog box.



The list box located at the left shows the available variables. The drop down list at the top of this box allows you to display only specific types of variables (such as numeric, string, date, etc.) or the variables belonging to a user defined set of variables. The Independent and Dependent list boxes located on the right of the dialog contain the variables that will be treated as independent and dependent. By default, the variable names are displayed in the same order as they appear in the data file. You may also use the Sorted check box to display the variable names in alphabetical order.


To assign variables as dependent or independent

Select the variables in the Variable list box and move them into the Independent or Dependent box by clicking on the  button next to the proper list box.

While most statistical analyses in SIMSTAT require you to distinguish between dependent and independent variables, some analyses do not require such a distinction. This is the case for several commands involving a single variable per analysis such as LIST, DESCRIPTIVE, FREQUENCY, TIME-SERIES, BINOMIAL, CHI-SQUARE TEST, and RUNS TEST. With those commands, variable assignments to the dependent or independent list boxes are simply ignored and the procedure is applied successively to all the independent variables followed by all the variables assigned to the dependent list box.


The RELIABILITY, the ITEM ANALYSIS and all multivariate analyses available from the OTHER submenu (i.e. Factor analysis, cluster analysis, etc.) also do not make a distinction between dependent and independent variables, and will include all variables in both lists in a single analysis. However, when performing a split-half reliability test using the RELIABILITY command, the two list boxes are used to determine which variables will be included in each versions of the instruments.


To remove variables from the dependent or independent list.

Select the variables you want to remove and click on the  button. To quickly remove a single variable from the list, double click on the variable name.

To apply an integer weight to your cases

SIMSTAT allows you to select a variable that will be used to weight the cases. When SIMSTAT reads a case, the value of the weighting variable for this case is truncated to an integer. This integer value specifies how many times the case will be duplicated. If the value is less than 1, the case is excluded from the analysis.

To assign a variable as a weighting variable, highlight in the variable list the variable you want to use as a weighting variable, and click on the  button next to the Weight box.

To remove the weighting variable, select this variable, and click on the  button next to the Weight box.

Binomial test

The BINOMIAL TEST allows you to assess whether the observed number of cases in a dichotomous variable is the same as that expected from a specified binomial distribution. The dialog box allows you to specify whether the comparison will be made on the observations above or below the mean, the median or a user-specified cutoff value, or to select observations equal to two values. It also allows specification of the test proportion.

OPTIONS

Cutoff point - This list box allows you to specify how the values of a variable will be dichotomized. The cutoff point used can be either the mean, the median or a user-specified value. All observations falling below the cutoff point form one group, and all observations equal to or above the cutoff point form the other group. The value mode can also be used to restrict the analysis to two groups defined by distinct values.

Values - This option is used only if you have selected **Value** as a cutoff point. If only one number is provided, it is used as a cutoff point. All observations falling below this value form one group, and all observations equal to or above the cutoff point form the other group. If two numbers are specified, cases with values equal to the first number form one group and cases with values equal to the second number form the second group. Providing no values allows the analysis of dichotomous variables.

Proportion - This option allows the specification of the test proportion. This value must lie between 0 and 1.

Sample output of a binomial test analysis

```

BINOMIAL TEST: AGGRESS  Nb of aggressive behaviors

      Mean =      27.5424
      Proportion =      .5000

      Cases

      37  Lt Mean      Observed proportion =  .6271
      22  Ge Mean
      ----
      59

      Z =  1.8226      2-tailed P =  .0684
  
```

Bootstrap analysis

The BOOTSTRAP submenu gives you access to an innovative and extremely powerful statistical technique called bootstrap simulation. This technique, developed by Efron (Efron, 1981; Diaconis & Efron, 1983) can be used to assess various properties of statistical estimators such as their accuracy, their sampling variability, etc. Typical applications include the computation of nonparametric estimates of sampling distributions, the assessment of the stability of statistical estimators, and the construction of nonparametric confidence intervals. SIMSTAT also allows the computation of nonparametric power estimates and Type I error rates for various estimators.

The following section provides a short non-technical introduction to the bootstrap technique followed by a description of SIMSTAT's particular implementation of bootstrapping methodology. Potential applications for researchers, statistical consultants, and for students and teachers of statistics are also presented. For further information about bootstrap methods and its applications, you can read the articles of Efron and his colleagues (Diaconis & Efron, 1983; Efron, 1981; Efron & Gong, 1983; Efron & Tibshirani, 1993). Wasserman and Bockenhold (1989) also provide an excellent introduction to bootstrap methodology, while Stine (1989) offers a comprehensive presentation of its potential applications.

What is bootstrap simulation ?

Bootstrap simulation is a resampling technique whereby initial sample subjects are treated as if they constitute the population under study. By replicating those data an infinite number of times, we then draw at random from that population a large number of samples, each the same size as the original sample. By computing a statistical estimator of interest (such as a mean or a correlation between two variables) for every bootstrap sample, this resampling procedure recreates an empirical sampling distribution of this estimator.

The main advantage of such a procedure is that the sampling distribution is not mathematically estimated but empirically reconstructed based on all the original characteristics of the data. So, it automatically takes into account distribution properties that are generally considered as contaminating factors, such as skewness, ceiling effects, outliers, etc. This feature makes bootstrap estimations adequate even when data are not normally distributed. In fact, bootstrapping can even be used to describe the sampling distribution of estimators for which sampling properties are unknown or unavailable.

SIMSTAT implementation of bootstapping

SIMSTAT provides automatic bootstrap analysis for seven descriptive estimators of a single variable and almost thirty estimators involving two variables. Those estimators are:

One variable estimators:

- Mean
- Median
- Variance
- Standard deviation
- Standard error
- Skewness
- Kurtosis

Two variables estimators:

- Kendall's tau-a and b
- Kendall-Stuart's tau-c
- Symmetric and asymmetric Somers' d
- Goodman-Kruskal's gamma
- Student's t and F
- Pearson's r
- Spearman's r_s
- Regression slope and intercept
- Mann-Whitney's U
- Wilcoxon's W
- Difference between means
- Difference between variances
- Sign test
- Kruskal-Wallis ANOVA
- Median test
- Percentage of agreement
- Cohen's kappa
- Scott's pi
- Krippendorff's R and \bar{r}
- Free marginal (nominal and ordinal levels)

The number of bootstrap samples for a single analysis can range from 10 to 100,000. The output of a simulation analysis can consist of various results, including descriptive statistics, frequency tables, histograms and percentile tables. The program also computes bootstrap confidence intervals. For estimators which can be tested for significance, SIMSTAT also displays nonparametric power estimates for up to four alpha levels. Power estimation with the bootstrap technique is straightforward: while performing bootstrap on a given data set, the proportion of redrawn samples that led to a statistically significant estimator (at some given alpha level) is computed and used as a power estimate. In addition to simulation results, the program displays the value of the seed used to initialize the random number generator. This value may then be used to regenerate the same data at a later time, or to compare various estimators using the same bootstrap samples.

The FULL ANALYSIS bootstrap command also offers the possibility to compute almost any available statistical analysis on successive bootstrap samples and displays the entire results of those analyses.

EXTENSIONS TO BOOTSTRAP

To achieve an even greater range of potential application, SIMSTAT implements two extensions to standard bootstrap simulation.

Variable sample size

Typically a bootstrap simulation randomly draws samples the same size as the original one. SIMSTAT offers the possibility to modify the dimension of the bootstrap samples, thus allowing users to compare estimator distributions obtained from using different sample sizes. You can set bootstrap simulations involving sample sizes that range from 1 to 100,000 observations.

Random sampling

Another aspect of bootstrapping is that it rests on the assumption that the original sample is representative of the population. SIMSTAT offers a modified bootstrap sampling process that rests on the null assumption that there is no difference or relation between variables in the population. While in bootstrap sampling the drawing is achieved by drawing vectors of scores for a particular subject, the RANDOM SAMPLING procedure draws individual subject scores for each variable independently. Consequently, while a standard bootstrap simulation on a correlation between two variables would yield coefficients that fluctuate around the correlation that exists in the original sample, the RANDOM SAMPLING procedure would produce correlations that vary around a null correlation. In this procedure, the proportion of redrawn samples that lead to a statistically significant estimator at a given alpha level is used to assess the Type I error rate.

Possible applications

We have already seen that standard bootstrap resampling can be used to obtain various measures of sampling variability such as nonparametric confidence intervals. The ability to alter the bootstrap sample size and to replicate the condition of the null hypothesis makes possible numerous new applications. The following paragraphs give some examples of such applications.

Research planning - Power estimation

The possibility of comparing various estimator distributions obtained for different sample sizes can prove useful in planning research by allowing the researcher to determine the sample size needed to achieve a desired precision level. It can also be used for power estimation, allowing comparison of the power attained using various estimators and/or sample sizes. Researchers thus have an empirical basis for choosing between two different statistical strategies. In addition, unlike standard approaches to power estimation, which rely on numerous assumptions, including normal data distributions, bootstrap power estimates make no distribution assumptions.

Teaching Tool

As a teaching tool, bootstrap simulation would be effective in illustrating to new statistics students concepts such as sampling theory or the central limit theorem. It would provide a simulation of the sampling process of an experiment, allowing the students to visualize the sampling variability of given estimators. By increasing or decreasing sample size, the student can observe how these changes affect the variability of estimators or the statistical power of an experiment. Additionally, bootstrap would be effective in demonstrating how outliers can affect estimation and how data transformation can improve population estimates.

Monte Carlo investigations

Bootstrap might also be handy for the researcher interested in studying the effect of violation of the normality assumption on some estimators by allowing the evaluation of the Type I and Type II (statistical power) error rate of a test. While Monte Carlo simulations usually analyze data generated by assumed mathematical functions, bootstrap simulation provides a direct assessment of sample distributions from data supplied by the researcher. By performing simulations on data distributions more representative of real-world data, bootstrap may be a more appropriate evaluation of statistical robustness.

The dialog boxes of the ONE VARIABLE, TWO VARIABLES, and RANDOM SAMPLING are very similar and have comparable options. The following section presents those options and provides an indication when an option is specific to an analysis.

OPTIONS

RESAMPLING PAGE

Estimator - This option displays a drop down list from which you can choose a specific measure that will be computed on each bootstrap sample. When the ONE VARIABLE command is chosen, you can select from a list of seven estimators (see above), while the TWO VARIABLES and the RANDOM SAMPLING command offer a choice between 28 estimators.

Values - Some estimators performed on independent samples require the specification of two values of the grouping (independent) variable. Those two values will be used to define two groups or will be treated as minimum and maximum values of the grouping under consideration (Option only available for ONE VARIABLE and RANDOM SAMPLING commands).

Number of samples - This option allows you to choose the number of times resampling is carried out. This number can range from 10 to 100,000.

Initial seed value - SIMSTAT automatically provides a “seed” value for the random number generator that drives the simulation analysis. Alternatively, the Initial Seed Value option can be used to specify a seed value. To instruct SIMSTAT to randomly

choose a new seed value, click on the **Shuffle** button. SIMSTAT outputs the seed value with the simulation results. This value can be used to regenerate the same data at a later time, or to compare various estimators using the same bootstrap samples.

Same as original - When this option is enabled the size of each bootstrap sample drawn from the original data is automatically adjusted to the size of this original sample. When disabled, you can use the **Size** option to specify how large each bootstrap sample will be. The sample size can range from 1 to 100,000 cases.

OUTPUT PAGE

Descriptive statistics - Enable this option to obtain a table of various descriptive statistics. This table includes the following statistics: mean, standard error of the mean, sum, mode, median, standard deviation, variance, skewness, kurtosis, minimum, maximum, and range.

Confidence intervals - The Confidence Intervals option allows you to display the 90%, 95%, and 99% confidence intervals of the coefficient. The program also displays bias-corrected confidence intervals that apply a correction to the interval to rectify situations where there is too much imbalance in the proportion of bootstrap estimates falling on each side of the observed statistic (median biased). The **Width** option allows you to define a fourth interval.

Percentile table - This option displays a percentile table whereby the number of cases in the sample is divided into equal categories. These categories indicate the percentage of cases that fall below the corresponding value of the variable. The number of categories computed in this table is determined in the **Number** option and is displayed with the corresponding variable values. For example, if you set this option to 4, the program will rank all valid cases, divide them into four equal groups, and then display the values that delimit the 25th (lower quartile), 50th (median) and 75th (upper quartile) percentiles.

Histogram - This option produces a graphic display of the distribution of a numeric variable. When this option is activated, the program first separates the values into non-overlapping intervals of equal width, and then plots bars that represent the relative frequencies of each interval. The **Number of bars** option allows you to specify how many bars (or intervals) to be plotted. You can also superimpose a **Normal curve** on the histogram, and visually assess how close your data distribution is to normal.

Type I error rate - This option gives the proportion of samples that have reached a specified alpha level. The standard display includes the proportions for 0.10, 0.05 and 0.01 alpha levels. The **Alpha** option lets you specify a fourth Alpha level (option only available for the RANDOM SAMPLING command).

Power estimate - The Power Estimate option gives the proportion of samples that have reached a specified alpha level. The standard display includes the proportions for 0.10, 0.05 and 0.01 alpha levels. The **Alpha** option lets you specify a fourth Alpha level (only available for the TWO VARIABLES command).

Full analysis bootstrap analysis

The FULL ANALYSIS bootstrap procedure allows one to perform various statistical analyses such as frequency analysis, crosstabulation or multiple regression on bootstrap samples. The program provides a complete analysis for each bootstrap sample. Specific statistics can then be extracted from the listing file with the use of a text editor and then stored in a new data file for further analysis.

The dialog box allows you to choose which analysis to perform, the sample size and the number of samples to be taken from the original sample. A second dialog box allows you to control how the analysis is to be performed and what statistics are to be printed.

OPTIONS

Analysis - Choosing this option evokes a drop down list from which you can choose a specific analysis to be performed on each bootstrap subsample.

Same as original - When this option is enabled the size of each bootstrap sample drawn from the original data is automatically adjusted to the size of this original sample. When disabled, you can use the **Size** option to specify how large each bootstrap sample will be. The sample size can range from 1 to 100,000 cases.

Number of samples - This option allows you to choose the number of times resampling is carried out on the data. This number can range from 1 to 32,000.

Seed - SIMSTAT automatically provides a “seed” value for the random number generator that drives the simulation analysis. Alternatively, the Initial Seed Value option can be used to specify a seed value. To instruct SIMSTAT to randomly choose a new seed value, click on the **Shuffle** button. SIMSTAT outputs the seed value with the simulation results. This value can be used to regenerate the same data at a later time, or to compare various estimators using the same bootstrap samples.

Breakdown analysis

The BREAKDOWN procedure computes descriptive statistics for sub-groups of the sample. This procedure can display a single line of statistics including the means, standard deviations, minimum and maximum values for the dependent variable (Y) within groups defined by the values of the independent variables (X). More detailed statistics can also be obtained for each group and for the entire sample. The dialog box also allows you to restrict the number of groups to a specified range and to obtain a multiple box-&-whisker plot that can be used to compare the distribution of the independent variable among several sub-groups.

OPTIONS

Range of X - This section requests two values that will be treated as minimum and maximum values of the grouping (or independent) variable under consideration. Each discrete value of the independent variable that falls within this range defines a distinct group. If those fields are left blank, all values of the independent variable will be included.

Statistics - When set to Brief, the mean, standard deviation, minimum and maximum values, and the number of cases are displayed on a single line. To obtain additional statistics, such as the skewness, kurtosis, mode, median, etc., set this option to Detailed.

Box-&-Whisker plot - The Box-&-Whisker Plot option allows you to obtain a multiple box-&-whisker plot that can be used to compare the distribution of the dependent variable among several sub-groups.

Sample output of a breakdown analysis

BREAKDOWN						
HOURSTV		Hours per week spent watching TV				
With SIBLING		Number of siblings				
Variable		Mean	Std Dev	Minimum	Maximum	Cases
Entire Population		15.02	3.76	5.50	23.00	59
SIBLING	.00	13.98	3.99	5.50	22.50	29
SIBLING	1.00	15.92	3.30	10.50	23.00	24
SIBLING	2.00	16.42	3.47	12.50	22.50	6

Correlations

The CORRELATION command produces a matrix of Pearson product-moment correlation coefficients for all pairs of dependent and independent variables. You can request either exact probabilities for the coefficients or a display of asterisks that indicate the probability levels attained. You can also tell the program to calculate probabilities using one- (directional) or two-tailed tests (non-directional) and to display cross-product deviation and covariance tables.

OPTIONS

Type of matrix - When set to X vs Y, the correlation matrix displays correlations between all variables assigned as independent against all variables assigned as dependent. The square matrix option produces a matrix displaying correlations between all selected variables, without taking into account whether they were selected as dependent or independent.

Confidence intervals - The confidence intervals option allows you to display confidence intervals of the correlations. The interval **Width** is expressed as a percentage between 1 and 99%.

Display probabilities values - This option determines the content of the correlation matrix. When disabled, the display includes up to 3 asterisks (*) to indicate the significance level attained for each correlation coefficient. If you enable this option, the program prints a matrix including the number of cases used to compute each coefficient and the estimated probability of the correlation.

Significance test - This option specifies whether the probability of the correlations is based on a one-tailed (directional) or two-tailed (non-directional) test.

Missing values - The Missing Values option allows you to specify whether you want to exclude cases with missing values by either PAIRWISE or LISTWISE deletion. If you select pairwise deletion, a case is excluded if it has a missing value on either of the two variables used to compute a given correlation coefficient. However, this case can be included in the computation of other coefficients. Listwise deletion excludes cases containing missing data from the computation of all the correlations included in the matrix.

Cross-product covariance matrix - This option displays cross-product deviation and covariance tables for the data.

Sample output of a correlation matrix

CORRELATION MATRIX: Pearson					
Variables	Cases	Cross-Prod Dev	Variance-Covar		
SEX	AGE	59	-4.2712	-.0736	
SEX	SIBLING	59	-1.3051	-.0225	
SEX	HOURLTV	59	-3.5085	-.0605	
SEX	AGGRESS	59	-265.2712	-4.5736	
AGE	SIBLING	59	-4.5254	-.0780	
AGE	HOURLTV	59	23.9576	.4131	
AGE	AGGRESS	59	399.6441	6.8904	
SIBLING	HOURLTV	59	38.3898	.6619	
SIBLING	AGGRESS	59	192.4746	3.3185	
HOURLTV	AGGRESS	59	1054.9576	18.1889	
	SEX	AGE	SIBLING	HOURLTV	AGGRESS
SEX	1.0000	-.0988	-.0666	-.0319	-.5471
	(59)	(59)	(59)	(59)	(59)
		P= .228	P= .308	P= .405	P= .000
AGE	-.0988	1.0000	-.0788	.0744	.2813
	(59)	(59)	(59)	(59)	(59)
	P= .228		P= .276	P= .288	P= .015
SIBLING	-.0666	-.0788	1.0000	.2630	.2988
	(59)	(59)	(59)	(59)	(59)
	P= .308	P= .276		P= .022	P= .011
HOURLTV	-.0319	.0744	.2630	1.0000	.2921
	(59)	(59)	(59)	(59)	(59)
	P= .405	P= .288	P= .022		P= .012
AGGRESS	-.5471	.2813	.2988	.2921	1.0000
	(59)	(59)	(59)	(59)	(59)
	P= .000	P= .015	P= .011	P= .012	
(Coefficient / (Case) / Probability 1-tailed)					

Crosstabs

The CROSSTAB command produces a standard contingency table for two variables where rows represent the dependent variable values and columns represent the independent variable values. The dialog box allows you to include various statistics in the table (count, row, column, or total percentage, etc.) and obtain various measures of association for nominal, ordinal and interval levels of measurement.

OPTIONS

TABLE PAGE

Contingency table - This option requests the output of a contingency table.

Sort by - Use this option to tell the program whether rows and columns of the table should be sorted by the values of the variable or by frequency.

Type - The type option allows you to specify whether rows and columns of the contingency table should be sorted in ascending or descending order.

Table content group - The Table Content group box allows you to request other statistics in addition to frequencies to be included in the cells of the table. To obtain the desired statistics, enable the corresponding check boxes:

- Row percentages
- Column percentages
- Total percentages
- Expected frequencies
- Chi-square residuals
- Standardized chi-square residuals

When performing multiple response crosstab analyses, an additional option allows you to specify whether the percentage will be based on the total number of responses, or on the number of respondents.

STATISTICS PAGE

The Statistics page allows you to specify which statistics should be computed on the table. You can specify as many statistics as needed in a single analysis.

Nominal level statistics

- Chi-Square and likelihood ratio statistics
- Phi for 2 x 2 tables or Cramer's V for larger tables
- Contingency coefficient

Ordinal level statistics

- Spearman's Rho
- Kendall's tau-b
- Kendall-Stuart's tau-c
- Goodman-Kruskal's gamma
- Somers' d

Continuous or interval level statistic

- Pearson's R

CHART PAGE

Barchart - If this option is activated, the program displays a 2 dimensions barchart that provides a graphical presentation of the relationship between the dependent and the independent variable.

Type - The Type option offers a choice between 4 different bivariate barcharts to display:

- In a **clustered** barchart, the bars representing the response frequency for each category of the independent variable are placed side by side.
- When the **overlayed** barchart is chosen, the bars are displayed on a 3 axis plane where bars representing the frequency of each category of the independent variable are placed on different rows. This chart type is available only in 3-D view. While this kind of chart is very popular, we strongly recommend against its use, since it is virtually impossible to determine the exact heights of the bars or compare the heights of bars located on different rows.
- In a **stacked** barchart, the bars representing the frequency of each category of the dependent variable are stacked on top of each other.
- The **100% bars** type is similar to the stacked barchart in that the bars for each category of the dependent variable are stacked on top of each other. However, like a pie chart, each bar represents the proportion of a category of the independent variable from the total number of observations in a specific category of the dependent variable. This type of barchart is especially useful if you want to

compare the proportions between different categories of the dependent variables rather than the absolute frequency.

Perspective- This option allows you to specify whether the barchart should be displayed in 2 or 3 dimensions.

A sample output of crosstabulation analysis

```

CROSSTAB

      SIBLING   Number of siblings
by SEX      Sex of the child

SIBLING->
      Count
      Row Pct
      Col Pct
SEX
-----
      .00      1.00      2.00      Total
Male
  1.00      13      13      3      29
      44.8      44.8      10.3      49.2
      44.8      54.2      50.0
Female
  2.00      16      11      3      30
      53.3      36.7      10.0      50.8
      55.2      45.8      50.0
Column
Total      29      24      6      59
      49.2      40.7      10.2      100.0

Chi-Square      Value      D.F.      Significance
-----
Pearson      .4602      2      .7945
Likelihood ratio      .4608      2      .7942

Smallest expected frequency = 2.949
Cells with expected frequency less than 5 = 2 of 6 (33.3%)

Statistic      Value      Significance
-----
Cramer's V      .08832
Contingency Coefficient      .08797
Pearson's correlation      -.06661      .6162
Spearman's correlation      -.07507      .5720
Kendall's Tau-b      -.07240      .5675
Kendall's Tau-c      -.07814      .5675
Gamma      -.13333
Somers' D :
  Symetric      -.07219
  With SIBLING dependent      -.07816
  With SEX dependent      -.06706
    
```

Descriptives

DESCRIPTIVES immediately computes univariate descriptive statistics for any variable assigned as a dependent or independent variable. Display includes mean, standard deviation, minimum, and maximum values for each variable. To obtain other descriptive statistics such as the skewness, kurtosis, mode, median, etc., use the FREQUENCY command.

Sample output of descriptive analysis

DESCRIPTIVE						
Variable	Mean	Std Dev	Minimum	Maximum	N	Label
SEX	1.51	.50	1.00	2.00	59	Sex of the child
AGE	9.54	1.48	6.00	11.00	59	Age of the child
SIBLING	.61	.67	.00	2.00	59	Number of siblings
HOURSTV	15.02	3.76	5.50	23.00	59	Hours per week spen
AGGRESS	27.54	16.58	.00	66.00	59	Nb of agresive beh

Factor analysis

EASY FACTOR ANALYSIS v3.0 performs two common types of factor analysis: Principal components analysis and image covariance factor analysis. The program has a good selection of features, such as variable criteria to limit factoring, varimax rotation, factor scores for both rotated and unrotated solutions, and flexible output. The program can also handle up to 100 variables, and an extremely large number of cases. (For more detailed information on factor analysis and the various statistics computed by EFA, see EFA user's manual).

NOTE: EFA v3.0 is an add-on module written by Dr Darren Fuerst and sold separately. To get more information on this module or to order a copy contact Provalis Research.

OPTIONS

Type of Analysis - Sets the type of factor analysis to use. The available types are Principal Components and Image Covariance.

Varimax Rotation - Set this option if you would like to rotate your factors using varimax orthogonal rotation.

Number of Factors - Sets the maximum number of factors to retain in subsequent analyses (i.e., rotation and scoring). The default is the number of variables in the data set.

Minimum Eigenvalue - Sets the minimum permissible eigenvalue for a factor to be retained in subsequent analyses (i.e., rotation and scoring). The number of factors to retain and minimum eigenvalue criteria are evaluated in an either/or fashion; that is, if either criterion is met, the number of factors retained is cut off at that point.

Descriptive statistics - If you would like means and standard deviations for your selected variables in the output, enable this option. Note that the standard deviations printed by EFA are *unadjusted* (i.e., the sum of the squared deviations is divided by n , rather than $n-1$). To obtain adjusted standard deviations or more detailed descriptive statistics see the DESCRIPTIVES or the FREQUENCY commands.

Correlation matrix - When toggled on, the correlation matrix for the variables in the data set will be included in the output.

Scree plot - If you would like a scree plot of the eigenvalues in the output, select this option.

Unrotated factor solution - When enabled, the results of the unrotated factor solution will be included in the output. You can turn this off if you're running multiple analyses on the same data and do not need to have this information repeated in the output.

Rotate factor solution - When enabled, the factors are rotated to the varimax criterion, and the results of this analysis are included in the output. You may wish to turn this option off during the initial analysis of a very large data set, when you're most interested in determining the number of factors in your data, as rotation of a large number of factors can be relatively time consuming.

Factor scoring weights - When enabled, EFA will calculate factor scores for both the rotated and unrotated solutions. The factor scoring coefficients will be printed in the output. By default, factor scores are not calculated, as they are not always needed, are relatively time consuming to calculate, and are usually not calculated until a final factor solution has been arrived at.

Display factor scores - If you would like the factor scores for your subjects to be printed in the output, toggle this option on. Be warned that including the factor scores in the output can increase its size dramatically.

Sample output of a factor analysis

```

FACTOR ANALYSIS
                                Easy Factor Analysis 3.0
                                (c) 1995 Darren Fuerst, All Rights Reserved

Analysis Parameters
-----
Path                .\
File                SIM2EFA
Analysis Type       Principal Components
#Factors            10
Min Eigenvalue      1.00000
Max Iterations      30
Rotate              Yes
Score               Yes

Analysis Log
-----
Analysis began 5/20 1996 at 20:58:31.
The raw data file has 12 variables and 132 observations.
Correlation matrix created.
Factor extraction complete.
NOTE: Trace = 12.000, with 63.46% of the total trace extracted by 3 factors.
Varimax rotation complete.
Unrotated factor scores calculated.
Rotated factor scores calculated.
Analysis ended 5/20 1996 at 20:58:32.

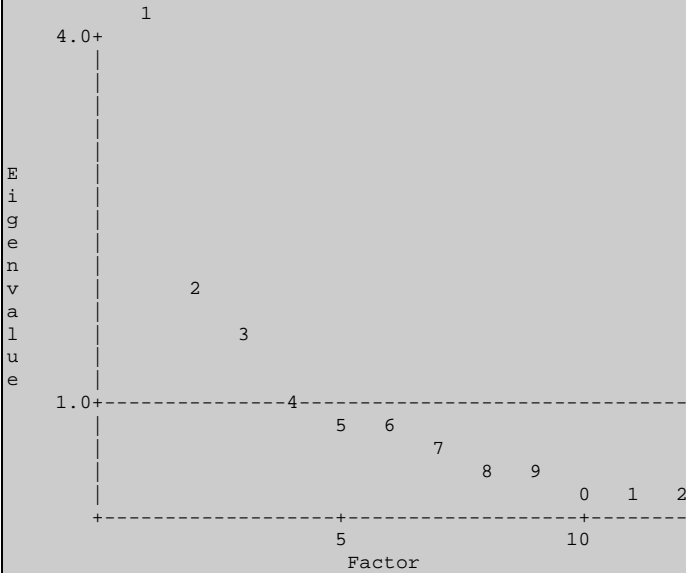
```

Correlation Matrix

	ACHIEVE	INTSCREE	DEVELOP	SOMAT	DEPRESS	FAMILY
ACHIEVE	1.0000	0.3386	0.6610	-0.0296	0.0667	0.2265
INTSCREE	0.3386	1.0000	0.4172	-0.0738	0.0448	-0.0262
DEVELOP	0.6610	0.4172	1.0000	0.0126	0.1977	0.0944
SOMAT	-0.0296	-0.0738	0.0126	1.0000	0.3482	0.1967
DEPRESS	0.0667	0.0448	0.1977	0.3482	1.0000	0.4195
FAMILY	0.2265	-0.0262	0.0944	0.1967	0.4195	1.0000
DELINQ	0.1275	0.0981	0.1755	0.2681	0.3936	0.2691
WITHDRAW	0.1624	0.0044	0.2336	0.1540	0.6755	0.3148
ANXIETY	-0.0330	0.0476	0.1030	0.2782	0.8033	0.3570
PSYCHOSI	0.2103	0.1166	0.3907	0.3255	0.7041	0.3182
HYPERACT	0.0424	0.0004	0.0276	0.1281	0.0876	0.1538
SOCSKILL	0.1174	0.1208	0.2713	0.1700	0.6189	0.3127

	DELINQ	WITHDRAW	ANXIETY	PSYCHOSI	HYPERACT	SOCSKILL
ACHIEVE	0.1275	0.1624	-0.0330	0.2103	0.0424	0.1174
INTSCREE	0.0981	0.0044	0.0476	0.1166	0.0004	0.1208
DEVELOP	0.1755	0.2336	0.1030	0.3907	0.0276	0.2713
SOMAT	0.2681	0.1540	0.2782	0.3255	0.1281	0.1700
DEPRESS	0.3936	0.6755	0.8033	0.7041	0.0876	0.6189
FAMILY	0.2691	0.3148	0.3570	0.3182	0.1538	0.3127
DELINQ	1.0000	0.2230	0.2349	0.4754	0.4673	0.5694
WITHDRAW	0.2230	1.0000	0.4263	0.6047	-0.1485	0.4500
ANXIETY	0.2349	0.4263	1.0000	0.4459	-0.0036	0.3962
PSYCHOSI	0.4754	0.6047	0.4459	1.0000	0.1780	0.6615
HYPERACT	0.4673	-0.1485	-0.0036	0.1780	1.0000	0.4163
SOCSKILL	0.5694	0.4500	0.3962	0.6615	0.4163	1.0000

Unrotated Solution: Scree Plot



Unrotated Solution: Eigenvalues

Factor1 Factor2 Factor3

Eigenvalue	4.2398	1.8856	1.4895			
Difference	3.4583	1.9376	0.0000			
%Trace	35.3320	15.7136	12.4124			
Cumulative	35.3320	51.0456	63.4580			
Unrotated Solution: Factor Pattern						
	Factor1	Factor2	Factor3			
DEPRESS	0.8744	-0.2521	-0.2444			
PSYCHOSI	0.8480	0.0205	-0.0343			
SOCKSKILL	0.7884	-0.0258	0.2806			
WITHDRAW	0.6856	-0.0777	-0.4527			
ANXIETY	0.6715	-0.3092	-0.3189			
DELINQ	0.6229	-0.0167	0.5434			
FAMILY	0.5343	-0.0892	-0.0131			
SOMAT	0.4008	-0.3056	0.0962			
ACHIEVE	0.2966	0.7815	-0.0520			
DEVELOP	0.4286	0.7614	-0.0979			
INTSCREE	0.1745	0.6531	-0.0199			
HYPERACT	0.2780	-0.0258	0.8519			
Varimax Rotation: Percentage Communalities						
ACHIEVE	INTSCREE	DEVELOP	SOMAT	DEPRESS	FAMILY	DELINQ
70.1442	45.7394	77.3038	26.3218	88.7968	29.3585	68.3650
WITHDRAW	ANXIETY	PSYCHOSI	HYPERACT	SOCKSKILL		
68.0991	64.8155	72.0757	80.3728	70.1031		
Varimax Rotation: Percentages of trace						
Factor1	Factor2	Factor3				
29.9706	17.3131	16.1743				
Varimax Rotation: Factor Pattern						
	Factor1	Factor2	Factor3			
DEPRESS	0.9311	0.0315	0.1412			
ANXIETY	0.8015	-0.0751	-0.0066			
WITHDRAW	0.7984	0.1610	-0.1327			
PSYCHOSI	0.7437	0.2663	0.3111			
SOCKSKILL	0.5803	0.1785	0.5766			
FAMILY	0.4954	0.0696	0.2081			
SOMAT	0.4002	-0.1843	0.2628			
DEVELOP	0.1851	0.8579	0.0523			
ACHIEVE	0.0463	0.8353	0.0401			
INTSCREE	-0.0344	0.6750	0.0253			
HYPERACT	-0.0903	-0.0162	0.8918			
DELINQ	0.3293	0.1176	0.7493			
Unrotated Scoring Weights						
	Factor1	Factor2	Factor3			
ACHIEVE	0.0700	0.4145	-0.0349			
INTSCREE	0.0411	0.3464	-0.0134			
DEVELOP	0.1011	0.4038	-0.0657			
SOMAT	0.0945	-0.1620	0.0646			

DEPRESS	0.2062	-0.1337	-0.1641
FAMILY	0.1260	-0.0473	-0.0088
DELINQ	0.1469	-0.0089	0.3648
WITHDRAW	0.1617	-0.0412	-0.3039
ANXIETY	0.1584	-0.1640	-0.2141
PSYCHOSI	0.2000	0.0109	-0.0230
HYPERACT	0.0656	-0.0137	0.5720
SOCSKILL	0.1860	-0.0137	0.1884
Varimax Rotated Scoring Weights			
	Factor1	Factor2	Factor3
ACHIEVE	-0.0483	0.4185	-0.0208
INTSCREE	-0.0617	0.3434	-0.0100
DEVELOP	-0.0059	0.4198	-0.0359
SOMAT	0.1044	-0.1328	0.1040
DEPRESS	0.2840	-0.0545	-0.0608
FAMILY	0.1269	-0.0082	0.0450
DELINQ	-0.0151	0.0032	0.3931
WITHDRAW	0.2736	0.0327	-0.2104
ANXIETY	0.2714	-0.0929	-0.1246
PSYCHOSI	0.1796	0.0698	0.0595
HYPERACT	-0.1668	-0.0422	0.5496
SOCSKILL	0.0905	0.0246	0.2479

Frequencies

FREQUENCY performs frequency and descriptive analysis on all selected variables (dependent or independent). The dialog box allows you to display a table with frequency counts for each value of a variable, the percentage of the count over all cases and over valid cases only, and the cumulative percentage over all valid cases. It allows you to obtain a bar chart, a Pareto, or a pie-chart for numeric and string variables. When used with numeric variables, you can also obtain percentile tables, and descriptive statistics (mean, median, mode, standard deviation, variance, skewness, kurtosis, minimum, maximum, and range) for each variable as well as histograms, box-&-whisker plots, cumulative distribution charts, and normal probability plots.

OPTIONS

ANALYSIS TAB

Frequency table - This option requests the output of a frequency table. This table includes the frequency counts for each value of a variable, the percentage of the count over all cases and over valid cases only, and the cumulative percentage over all valid cases.

Sort by - Use this option to tell the program whether the frequency table should be sorted by the values of the variable or by order of frequency.

Type - The Type option allows you to specify whether the frequency table should be sorted in ascending or descending order.

Descriptive statistics - Use this option to obtain a table of various descriptive statistics. This table includes the following statistics: mean, standard error of the mean, sum, mode, median, standard deviation, variance, skewness, kurtosis, minimum, maximum, and range.

Confidence interval - This option prints a confidence interval around the mean. The **Width** option allows you to set the interval width expressed as a percentage. Its value should lie between 1% and 99%.

Percentile table - This option displays a percentile table whereby the number of cases in the sample is divided into equal categories. These categories indicate the percentage of cases that fall below the corresponding value of the variable. The number of categories computed in this table is determined in the **Number** option and is displayed with the corresponding variable values. For example, if you set this option to 4, the program will rank all valid cases, divide them into four equal groups, and then display the values that delimit the 25th (lower quartile), 50th (median) and 75th (upper quartile) percentiles.

CHARTS PAGE

Bar chart - If this option is activated, the program produces a bar chart that provides a graphical representation showing the relative frequencies of every value of a variable.

Pie chart - This option displays a pie chart where the relative frequency of each value is represented by a slice. Pie charts are appropriate when you want to compare individual values to other values and to the whole.

Pareto chart - This option allows you to obtain a bar graph that displays the categories of a variable, sorted in descending order of frequency, with a line chart above the bars to represent the cumulative percentages of the cases.

Box-&-Whisker - This option allows you to obtain a box-&-whisker plot that can be used to examine the distribution of the variable. It is especially useful to detect the presence of outliers and asymmetry in the data distribution. The box includes values that fall between the first and the third quartiles (about 50% of the values). The line in the middle of the box represents the median value, while the whiskers extend to the farthest observations within 1.5 times the interquartile range measured from the nearest quartiles. Values that are situated further than 1.5 times the interquartile range, but within 3 times this distance, are represented by the letter O (for outliers). Values farther than 3 times the interquartile range from the nearest quartile are represented by the letter X (for extreme).

Normal plot - This option displays a normal probability plot that allows you to evaluate whether the data are normally distributed. If the data follow a normal distribution, the data points will fall approximately along a straight line going from the lower left corner of the graph to the upper right corner.

Cumulative distribution - This option displays a cumulative distribution of frequencies.

Histogram - The Histogram option graphically displays the distribution of a numeric variable. When this option is activated, the program first separates the values into non-overlapping intervals of equal width, then plots bars that represent the relative frequencies of each interval. The **Number** option allows you to specify how many bars will be plotted. You can also assess how close the distribution is to a normal distribution by superimposing a normal curve on the histogram. To obtain such a curve, enable the **Normal curve** option.

OPTIONS PAGE

Suppress table - This option allows you to suppress the printing of frequency tables and nominal charts such as barcharts, piecharts, and Pareto charts when a variable has more values than the supplied cut-off value.

Sample output of a frequency analysis

FREQUENCIES: AGGRESS Nb of aggressive behavior					
	Value	Frequency	Percent	Valid Percent	Cum Percent
	.00	3	5.1	5.1	5.1
	2.00	1	1.7	1.7	6.8
	6.00	2	3.4	3.4	10.2
	8.00	1	1.7	1.7	11.9
	12.00	1	1.7	1.7	13.6
	14.00	8	13.6	13.6	27.1
	16.00	2	3.4	3.4	30.5
	18.00	1	1.7	1.7	32.2
	20.00	3	5.1	5.1	37.3
	24.00	6	10.2	10.2	47.5
	26.00	9	15.3	15.3	62.7
	30.00	2	3.4	3.4	66.1
	32.00	1	1.7	1.7	67.8
	34.00	2	3.4	3.4	71.2
	36.00	3	5.1	5.1	76.3
	46.00	8	13.6	13.6	89.8
	48.00	1	1.7	1.7	91.5
	53.00	1	1.7	1.7	93.2
	56.00	1	1.7	1.7	94.9
	66.00	3	5.1	5.1	100.0
	TOTAL	59	100.0	100.0	
Mean	28.542	Variance	274.839	Kurtosis	-.244
Median	27.000	Std Dev	16.578	S.E. Kurt.	.638
Mode	27.000	Std Err	2.158	Skewness	.531
Minimum	1.000	Sum	1684.000	S.E. Skew.	.319
Maximum	67.000	Range	66.000	Valid	59.000
95% Confidence Interval for the mean = [24.311 to 32.773]					
Percentile	Value	Percentile	Value	Percentile	Value
10.00	6.00	20.00	14.00	30.00	16.00
40.00	24.00	50.00	26.00	60.00	26.00
70.00	34.00	80.00	46.00	90.00	48.00

Friedman Test

The FRIEDMAN TEST is a procedure for testing whether two or more related samples have been drawn from the same population. The output displays the mean rank of each variable, the number of cases, chi-square, degrees of freedom and probability value.

Because the Friedman test is used to compare correlated samples, it does not really make a distinction between dependent or independent variables. The Variable Selection dialog box is used instead to specify which variables should be compared together. For example, assigning T1DEPRESS, T2DEPRESS, and T3DEPRESS to the Independent list box and T1AGGRESS, T2AGGRESS, and T3AGGRESS to the Dependent list box will result in two separate Friedman tests, the first one comparing the first 3 variables, and the second one comparing the other 3.

Sample output of a Friedman two-way anova for K related samples

FRIEDMAN TWO-WAY ANOVA			
Mean rank	Variable		
2.211	T1DEPRES		
2.118	T2DEPRES		
1.671	T3DEPRES		
Cases	Chi-square	D.F.	Significance
38	6.3289	2	.0422

GLM ANOVA/ANCOVA

GLM ANOVA/ANCOVA is a procedure which permits the researcher to examine the effects of one or more variables on a single continuous dependent variable. This procedure provides a way to test the equality of means in categories within a single variable or factor (main effects) as well as categories formed by combinations of two or more variables or factors (interaction effects). Analysis of covariance (ANCOVA) allows the comparison of the effect of categorical variables on the dependent variable while controlling for the effects of one or more quantitative variables (covariates). SIMSTAT's implementation of analysis of variance and covariance is based on the General Linear Model. Using a hierarchical regression analysis technique allows much greater flexibility than standard ANOVA/ANCOVA procedures by allowing one to freely combine quantitative and categorical factors and to statistically control for covariates which are either categorical or quantitative. The procedure can also be used to perform standard multiple regression problems that involve interaction terms. The GLM ANOVA/ANCOVA procedure also handles balanced and unbalanced ANOVA designs by providing automatic adjustment for unequal cell size. The dialog box allows you to display standard ANOVA/ANCOVA tables as well as various outputs usually found in ANOVA/ANCOVA or multiple regression analysis. Various adjustment methods for unequal cell sizes are provided, including a hierarchical strategy where the user can set the order of entry of each variable in the model.

OPTIONS

ANALYSIS PAGE

Show steps - This option requests the printing of various statistics at each step of the analysis. The information output at each step includes an ANOVA table, and a choice of statistics such as multiple regression coefficients, regression equation for the variables entered in the model, as well as cell adjusted means for nominal variables.

Multiple regression - This option displays the multiple regression coefficients, R square, adjusted R square, and the probability (significance) for the whole model (omnibus test).

Equation - When enabled, this option displays various statistics computed for the variables included in the model, including the regression coefficient (or B_i), its standard error and confidence limits, the standardized coefficient (or beta), whole, partial and semi-partial correlation coefficients, F ratio and its probability. SIMSTAT generates coded vectors to represent independent categorical variables (factors) and uses multiple regression techniques using those vectors in order to accomplish ANOVA or ANCOVA problems and obtain a regression equation that can be used for prediction. Effect coding is used to create the vectors such that for each vector created, cases of one group are assigned 1's while all other cases are assigned 0's with the exception of the cases of the

last group, which are coded as -1's. The following table illustrates the result of an effect coding where 3 vectors (X1, X2, and X3) are created in order to represent the various values contained in the GROUP variable.

GROUP	X1	X2	X3
1	1	0	0
1	1	0	0
2	0	1	0
2	0	1	0
3	0	0	1
3	0	0	1
4	-1	-1	-1
4	-1	-1	-1

The regression equation obtained with this method provides much useful information. When the sample sizes are equal and there is no covariate, the intercept is equal to the grand mean of the dependent variable. When the analysis involves unequal cell sizes, the intercept is equal to the unweighted mean, that is, to the average of all group means. Each coefficient associated with a coded vector represents the deviation from the grand mean for the group associated with this vector. The predicted score of a subject is obtained by adding to the grand mean the regression coefficient of the group to which the subject belongs. The specific effect of the last group can be obtained by computing the summation of all the b coefficients of the variables associated with the factor and inverting the sign of the result (or $-b_k$). The F and the significance value associated with the coefficients cannot be used to specify whether there are significant differences among the various groups but represent instead a test of the deviation of a group mean from the grand mean. When the analysis involves more than one factor, the mean of each cell can be obtained by substituting for each coded vector the proper code (i.e. 1, 0 or -1) in the regression equation. For instance, the predicted score of a subject with treatment combination A_i and B_j is equal to the sum of the intercept (grand mean), the effect of treatment i of factor A, the effect of treatment j of factor B, plus the effect associated with the interaction between those two treatments. When a quantitative variable (covariate) is included in the model, the predicted score can be obtained by adding to the result the value obtained by multiplying the observed value on this quantitative variable by its b coefficient.

Confidence interval - This option allows you to set the width of the confidence interval for the unstandardized regression coefficients. This interval width is expressed as a percentage and must lie between 1% and 99%.

Adjusted means - This option prints the predicted mean for each cell of all the nominal (categorical) variables entered in the model controlling for every quantitative variable (covariate) and/or interaction also in the model.

Test of change - This option allows the printing of a summary report of the changes in R square at each step of the analysis.

Adjustment method - This option allows you to choose between 3 types of adjustment for unbalanced designs. In the Regression model, all factors, covariates, and interactions are entered simultaneously; in the Nonexperimental approach covariates are entered first, followed by categorical factors and then by interactions; the Hierarchical approach allows you to specify the order in which each variable will be entered in the model. The following table shows the correspondence between those 3 methods and the terminology used by other sources:

SIMSTAT	OVERALL & SPIEGEL	SPSS/PC+	SAS
Regression	Method 1	Regression	Type III
Nonexperimental	Method 2	Classic Experimental	Type II
Hierarchical	Method 3	Hierarchical	Type I

DIAGNOSIS PAGE

Residual caseplot - This option allows the display of a casewise plot of standardized residuals, including the predicted, obtained, and residual values for all cases. This option is useful for identifying outliers (i.e., cases that are not well represented by the regression model). The **Outliers** option allows you to restrict the residual caseplot to those cases for which the absolute standardized value is greater than or equal to the specified value. The Outliers value can be set between 0 and 4 standard deviations.

Durbin-Watson statistic - This option tests for the presence of autocorrelation or serial correlation in the residuals. The larger the autocorrelation, the less reliable the results of the analysis.

Residual scatterplot - This procedure produces a bivariate scatterplot. The predicted value is plotted along the horizontal axis, and the standardized residuals on the vertical axis.

Residual normal plot - This option displays a normal probability plot of residual values that allows you to evaluate whether those residuals are normally distributed. If the residuals follow a normal distribution, the data points will fall approximately along a straight line going from the lower left corner of the graph to the upper right corner.

Save residuals - This option instructs SIMSTAT to save the predicted and residual values as new variables in the current data file. Creating those variables allows you to perform further analyses such as displaying scatterplots between each independent

variable and the predicted and residual values. The new variables are named PREDxxx and RESIDxxx where xxx stands for a serial number between 001 and 999 that corresponds to the number of analysis performed during a single command. This number is automatically reset to 001 after each command. To prevent overwriting those variables, you must rename them using the DEFINE VARIABLE command.

ORDER & INTERACTION PAGE

Type - This column allows you to specify whether the variable is nominal or quantitative. To toggle between these two types of variables, click on the appropriate radio button at the bottom of the grid.

Order - When a hierarchical approach is chosen (see Adjustment Method), this option allows you to specify the sequence of entry of the variables in the model. The order for each variable can lie between 1 and 6. All variables with the same number are entered at the same time, while variables that have lower numbers are entered before those that have higher numbers. Variables for which the order is set to 6 are entered at the same time as the interaction variables. You can enter the order at the keyboard or use the spin button located at the right edge of the grid to increase or decrease the order.

Interactions - This option allows you to specify which interactions are tested. For example, to test for interactions between factors A and B, and factors A, B and covariate C, enter:

Interactions: AB ABC

Sample output of a GLM ANOVA/ANCOVA analysis

```

GLM - ANALYSIS OF VARIANCE
Dependent Variable:  AGGRESS  Nb of aggressive behavior
Adjustment for unequal cells size:  Hierarchical
***** STEP 1 *****
Variable(s) entered on step 1:      AGE  Age of the child
Multiple Regression
      Multiple R      =  .2813          sig. of R =  .0309
      Multiple R Square =  .0791
      Adjusted R square =  .0630
Analysis of Variance

```

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Variables block #1	1261.136	1	1261.136	4.897	.031
AGE	1261.136	1	1261.136	4.897	.031
Explained	1261.136	1	1261.136	4.897	.031
Residual	14679.508	57	257.535		
Total	15940.644	58	274.839		

Regression equation						
Parameter	Estimate	Std Err	80% Confidence interval			
Intercept	-1.5700					
B1 : AGE	3.1556	1.4260	1.3071 to			5.0042
Parameter	Beta	Correl	S-Part	Partial	F ratio	Sig F
B1 : AGE	.2813	.2813	.2813	.2813	4.553	.0375
***** STEP 2 *****						
Variable(s) entered on step 2: SEX Sex of the child						
SIBLING Number of siblings						
Multiple Regression						
Multiple R	= .6958	sig. of R = .0000				
Multiple R Square	= .4842					
Adjusted R square	= .4460					
Analysis of Variance						
Source of Variation	Sum of Squares	DF	Mean Square	F	P	
Variables block #1	1261.136	1	1261.136	8.283	.006	
AGE	1261.136	1	1261.136	8.283	.006	
Variables block #2	6457.207	3	2152.402	14.136	.000	
SEX	4140.360	1	4140.360	27.192	.000	
SIBLING	2115.276	2	1057.638	6.946	.002	
Explained	7718.343	4	1929.586	12.673	.000	
Residual	8222.301	54	152.265			
Total	15940.644	58	274.839			
Regression equation						
	B	Std Err	80% Confidence interval			
Intercept	3.8458					
A1 : SEX = 1.00	8.4532	1.6211	6.3518 to			10.5546
B1 : AGE	3.0895	1.1107	1.6497 to			4.5293
C1 : SIBLING = 0.00	-7.3120	2.4352	-10.4688 to			-4.1553
C2 : SIBLING = 1.00	-5.8700	2.5095	-9.1230 to			-2.6170
Parameter	Beta	Correl	S-Part	Partial	F ratio	Sig F
A1 : SEX = 1.00	.5142	.5471	.5096	.5787	26.688	.0000
B1 : AGE	.2754	.2813	.2719	.3540	7.594	.0080
C1 : SIBLING = 0.00	-.2955	-.2988	-.2935	-.3782	8.849	.0044
C2 : SIBLING = 1.00	-.2302	-.1612	-.2286	-.3033	5.370	.0244

Cells adjusted means (adjusted for AGE)		
SEX	SIBLING	Adj. mean
1.0	.0	34.4681
1.0	1.0	35.9102
1.0	2.0	54.9622
2.0	.0	17.5617
2.0	1.0	19.0038
2.0	2.0	38.0558

***** STEP 3 *****

Interactions entered on step 3: SEX AGE

Multiple Regression

Multiple R = .6965 sig. of R = .0000

Multiple R Square = .4851

Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	F	P
Variables block #1	1261.136	1	1261.136	8.144	.006
AGE	1261.136	1	1261.136	8.144	.006
Variables block #2	6457.207	3	2152.402	13.900	.000
SEX	4140.360	1	4140.360	26.737	.000
SIBLING	2115.276	2	1057.638	6.830	.002
2-way interactions	15.036	1	15.036	.097	.756
SEX AGE	15.036	1	15.036	.097	.756
Explained	7733.379	5	1546.676	9.988	.000
Residual	8207.265	53	154.854		
Total	15940.644	58	274.839	Regression equation	

Parameter	B	Std Err	80% Confidence interval	
Intercept	4.3637			
A1 : SEX = 1.00	11.8408	10.9935	-2.4103 to	26.0918
B1 : AGE	3.0424	1.1303	1.5772 to	4.5076
C1 : SIBLING = 0.00	-7.4193	2.4798	-10.6340 to	-4.2047
C2 : SIBLING = 1.00	-5.7874	2.5445	-9.0859 to	-2.4889
A1*B1	-.3550	1.1394	-1.8320 to	1.1219

Parameter	Beta	Correl	S-Part	Partial	F ratio	Sig F
A1 : SEX = 1.00	.7203	.5471	.1062	.1464	1.160	.2863
B1 : AGE	.2712	.2813	.2653	.3468	7.245	.0095
C1 : SIBLING = 0.00	-.2998	-.2988	-.2949	-.3801	8.951	.0042
C2 : SIBLING = 1.00	-.2269	-.1612	-.2242	-.2982	5.173	.0270
A1*B1	-.2085	.5342	-.0307	-.0428	.097	.7566

Cells adjusted means (adjusted for AGE + Interactions)

SEX	SIBLING	Adj. mean
1.0	.0	34.4286
1.0	1.0	36.0605
1.0	2.0	55.0547
2.0	.0	17.5227
2.0	1.0	19.1546
2.0	2.0	38.1488

History of Changes in Multiple Regression

Step	Variable(s)	R Square	Rsq Chg	F Chg	D.F.	Prob.
1	AGE	.0791	.0791	4.897	1	.0309
2	SEX SIBLING	.4842	.4051	14.136	4	.0000
3	2-way interactions	.4851	.0009	.097	5	.7566

Standardized residual caseplot

Case #	Predicted	Obtained	Residual	-3.0	0.0	3.0
				:.....:.....:		
3	32.9710	67.0000	34.0290	.	.	* .
4	37.2903	67.0000	29.7097	.	.	* .
18	32.9710	7.0000	-25.9710	.	*	. .
22	58.9718	27.0000	-31.9718	.	*	. .
				:.....:.....:		
Case #	Predicted	Obtained	Residual	-3.0	0.0	3.0

Durbin-Watson test = 1.5870

Inter-raters analysis

Inter-rater agreement measures are used to assess the concordance in observed ratings of two judges at the same point in time. Such measures can also be used to assess the reliability of the ratings of a single judge at different points in time. The simplest measure of agreement for nominal level variables is the proportion of concordant ratings out of the total number of ratings made. Unfortunately, this measure often yields spuriously high values because it does not take into account chance agreements. Several adjustment techniques have been proposed in the literature to correct for the chance factor, three of which are available in the SIMSTAT program. The following are the assumptions made by each of these correction techniques: free marginal adjustment assumes either that all categories on a given scale have equal probability of being observed, or that the judges have not based their decisions on information about the distribution for their ratings. Scott's pi adjustment does not assume that all categories have equal probability of being observed, but does assume that the distributions of the categories observed by the judges are equal. Cohen's kappa adjustment does not assume that all categories have equal probability of being observed, nor that the distribution of the categories is equal for the two judges. It does, however, in computing the chance factor, take into account the differential tendencies of the judges.

SIMSTAT also offers three adjustments for ordinal level variables. These are similar to the previous measures except that they also take into account the ordinal nature of the scales by adjusting the weights assigned to various levels of agreement. They apply the same tree model of chance agreement used in the previous measures of nominal data. Free marginal adjustment for ordinal level variables also assumes that all categories on a given scale have equal probability of being observed. Krippendorff's R-bar adjustment is the ordinal extension of Scott's pi and assumes that the distributions of the categories are equal for the two sets of ratings. Krippendorff's r adjustment is the ordinal extension of Cohen's Kappa in that it adjusts for the differential rating tendencies of the judges in the computation of the chance factor.

OPTIONS

TABLE PAGE

Agreement table - This option requests the output of an agreement table.

Sort by - Use this option to tell the program whether the row and columns of the table should be sorted by the values of the variable or by frequency.

Type - The type option allows you to specify whether the rows and columns of the agreement table should be sorted in ascending or descending order.

Table content group - The Table Content group box allows you to request other statistics in addition to frequencies to be included in the cells of the table. You can obtain as many of the statistics as desired by enabling the corresponding check box:

- Row percentages
- Column percentages
- Total percentages
- Expected frequencies
- Chi-square residuals
- Standardized chi-square residuals

NOTE: The expected frequencies displayed in the inter-rater agreement tables do not necessarily correspond to the expected frequencies used in the above correction techniques. Rather, they correspond to the values used in the computation of chi-square statistics used in contingency tables. However, those values coincide with those used in the computation of Cohen's Kappa and Krippendorf's r.

STATISTICS PAGE

The STATISTICS page allows you to request various nominal and ordinal level measures of agreement:

- Percentage of agreement
- Cohen's kappa
- Scott's pi
- Free marginal (nominal)
- Krippendorf's r-bar
- Krippendorf's R
- Free marginal (ordinal)

CHART PAGE

Barchart - If this option is activated, the program displays a 2 dimensional barchart that provides a graphical presentation of the relationship between the dependent and the independent variable.

Type - The Type option offers a choice between 4 different ways of displaying a bivariate barchart:

- In a **clustered** barchart, the bars representing the frequency of each category of the independent variable are placed side by side.
- When the **overlayed** barchart is chosen, the bars are displayed in perspective on a 3 axis plane where bars representing the frequency of each category of the independent variable are placed on different rows. This chart type is only available in 3-D view. While this kind of chart is very popular, we strongly recommend

against its use, since it is virtually impossible to determine the exact heights of the bars or compare the heights of bars located on different rows.

- In a **stacked** barchart, the bars representing the frequency of each category of the dependent variable are stacked on top of each other.
- The **100% bars** type is similar to the stacked barchart in that the bars for each category of the dependent variable are stacked on top of each other. However, just like a pie chart, each bar represents the proportion of a category of the independent variable from the total number of observations in a specific category of the dependent variable. This type of barchart is especially useful if you want to compare the proportions between different categories of the dependent variables rather than the absolute frequency.

Perspective- This option allows you to specify whether the barchart should be displayed in 2 or 3 dimensions.

A sample output of inter-raters analysis

```

INTER-RATERS TABLE

      EVENT1J2  Level of stress - Judge #2
by EVENT1J1  Level of stress - Judge #1

EVENT1J2->
      Count      Low      Medium      High      Total
      Tot Pct
EVENT1J1 -----
Low      1.00      24      4      .0      28
      40.7      6.8
Medium   2.00      .0      22      2      24
      .0      37.3      3.4
High     3.00      1      2      4      7
      1.7      3.4      6.8
      Column      25      28      6      59
      Total      42.4      47.5      10.2      100.0

INTER-RATER AGREEMENT MEASURES

Nominal level

      Pct agreement = 84.7%
      Cohen's Kappa = 74.3%
      Scott's pi = 74.2%
      Free marginals = 77.1%

Ordinal level

      Krippendorff's r bar = 77.1%
      Krippendorff's R = 77.1%
      Free marginals = 84.7%
    
```

Item Analysis

STATITEM v1.0 performs classical item and test analysis for multichoice item questionnaires. It computes the percentage endorsement and the item-total correlations for correct and alternate responses. It also provides endorsement rates for various achievement levels and descriptive statistics on the total score such the mean, median, minimum, maximum, variance, skewness, kurtosis, etc., as well as Cronbach's alpha internal consistency and Ferguson's discrimination index. Descriptive statistics are also computed on percent correct and item total correlations. STATITEM is closely integrated with SIMSTAT and will perform analyses on data stored in each file format currently supported by SIMSTAT. The key responses used to specify correct responses may be stored in either the first record of the data file or in a key file.

NOTE: STATITEM v1.0 is an add-on module sold separately. To get more information on this module or to order a copy contact Provalis Research.

INTRODUCTION TO ITEM ANALYSIS

Usually, when one is developing a scale not all items will work as expected; Some may be confusing to the respondent, some may not tell us what we thought they would, others may be too easy or too hard to answer, and so on. Besides inspection and modification of wording to eliminate jargon terms, ambiguous formulations, double-barrelled questions, negatively stated items, etc., it is common practice to submit the items to a field test involving a sample of subjects and to apply several statistical techniques to the results in order to identify items that are not functioning as intended. These techniques allow one to improve the reliability and validity of a test by identifying items that should be eliminated, substituted or revised. In doing so, item analysis makes it possible to increase the overall quality of a scale while shortening it either by eliminating unsatisfactory items or by removing redundant ones (i.e., items that are equivalent to those that remain and thus provide no further information).

ASSESSING ITEM CONTRIBUTION TO INTERNAL CONSISTENCY

Because a scale is developed to measure an attribute, every question should tap an aspect of that attribute (e.g., attitude, behavior, knowledge) and, ideally, only that attribute. Since all items must reflect the attribute, they should share a common variance and should be related to the other items and, consequently, to the scale as a whole. The correlation between an item and the total score, omitting this item, can be considered as an indication of this relationship. Another desired attribute of any scale is its reliability, expressed as the stability of the score when the same instrument or an alternate form is administered to the same individuals. The Cronbach's alpha represents the most common index of the overall reliability or, more precisely, the internal consistency of a scale. The alpha coefficient is also an indicator of the scale precision and can be viewed as the theoretical maximum value a correlation with this scale can take.

Usually, the longer the scale, the higher the value of the alpha coefficient. Consequently, we may be tempted to always prefer a long version of a questionnaire to a shorter one. There are at least two problems with this solution. First, it is often desirable, for practical reasons to have a shorter test in order to reduce the administration time or cost. Second, it often occurs that some items, while positively correlated with the scale total, may reduce the overall reliability of the scale or may contribute only marginally to this reliability. The test developer needs to eliminate items without greatly reducing the scale's reliability. To achieve this, he/she needs to know the specific contribution of each item to the scale reliability index. The **Alpha if item deleted** represents such a measure. As the name suggests, it represents the Cronbach's alpha value we would obtain if the item was removed from the test. By successively eliminating items that reduce the Cronbach's alpha, or that contribute only slightly to the reliability index of the total score, it becomes possible to significantly reduce the number of items while maintaining an acceptable level of reliability.

MEASURING ITEM DIFFICULTY AND DISCRIMINATION

Many scales developed by psychologists and educational researchers are designed to differentiate among individuals at all levels of the measured construct, such as when a scale is built to differentiate subjects according to their level of achievement or ability, either for selection or classification purposes. Several statistics can be used to assess the discriminative quality of an item in a questionnaire. The most well known of these statistics is the difficulty index. Any response that is answered correctly or missed by all the examinees is useless since it will not allow one to differentiate between individuals. The item difficulty index, obtained by computing the percentage of subjects who responded correctly to the item, represents a crude indication of the item's capability to discriminate between examinees. From the perspective of the latent trait model, if we assume a perfect relationship between the ability of the examinees and the success on an item, then an item with a difficulty index of 50% would differentiate between those subjects who are positioned on the higher half of the ability scale from those in the lower half. In this ideal situation, a scale with all of its items at the same difficulty level of 50% would not be adequate since it would only differentiate between subjects in two groups and would not allow one to differentiate further among respondents in either of these groups. It would then be important to choose items with various levels of difficulty in order to be able to differentiate between all levels. However, such a perfect relationship between ability level and item response is seldom achieved, partly because items are seldom homogeneous. According to Henrysson (1971), when the level of correspondence between the item response and ability level - as measured by the item-total correlation - is moderate (between .30 and .40), then a scale composed of items with difficulty levels between 40% and 60% should provide an adequate test that will permit reliable discrimination between nearly all ability levels. If the average item-total correlation is higher, a wider range of difficulty levels is required.

When selecting items according to their difficulty level, two additional factors should be considered. While a mean difficulty level of 50% will usually maximize the total score variance, this value may need to be increased to take into account the fact that, with multiple choice items, a specific proportion of correct responses may be obtained simply by guessing. Also, the choice of an average difficulty level should take into account the fact that the scale may be constructed to differentiate between those at the lower or the higher ends of the scale. In such situations,

better discriminatory value will be achieved by selecting items with difficulty levels near the desired end.

However, the percentage of correct responses is not a sufficient condition to judge the quality of an item, since the number of correct responses should also be related to the level of the ability we want to measure. If those who answer correctly are low on the measured ability while the subjects with higher scores choose a wrong answer, or if the success on this item is the same for these two groups, then there may be a problem in the item formulation or the item may be measuring a different ability, unrelated to the one of interest. Several methods exist to assess the relationship between level of endorsement and the measured ability. A simple method consists of making sure that the percentage of correct responses to an item is higher for those who are high on the measured construct than for those who are low on this construct. In the absence of an external criterion, a second method uses the total score on all items as the discriminatory criterion. In this situation, we simply compute the total scores on the test for each subject and retain two distinct groups, each representing a specific proportion of the examinees (say 10, 25 or 50%) positioned at the upper and lower end of the ability scale. Once the upper and lower groups have been composed, we can obtain an index of discrimination for an item by computing the difference in the percentage of correct responses between the upper and the lower groups. This index (often designated as the D index) varies from -1.0 to 1.0, a positive value indicating that the item correctly discriminates according to the measured construct. The greater the difference, the better the item is able to discriminate between subjects. On the other hand, a negative value suggests a negative discrimination favoring those in the lower group, and is strong evidence for a problematic item. Ebel (1965) suggested the elimination or complete revision of items with a discrimination index less than .20 and the revision of those with an index value between .20 and .30.

While the D index is easily computed and interpreted, it suffers from a major drawback. When comparing only the two most extreme groups, much information is discarded such as the percentage of success in intermediate groups or among the subjects within a group. STATITEM offers several ways to examine the distribution of correct responses more closely. The program can provide detailed information about the distribution of correct responses by breaking down the total number of examinees into several groups (from 2 to 10) of similar size and displaying the percentage of correct responses for these groups. It is also possible to display even more detailed information on the success rate of an item for every value of the total score, allowing one to assess whether the item can provide adequate discrimination of subjects all along the scale. To simplify the examination of the relationship between the proportion of success to an item and the total score, STATITEM also provides a graphical representation of this relationship as an empirical item characteristic curve (also called an item-test regression curve) which displays, for each total score level, the percentage of persons who responded correctly. This curve allows one to examine the item's difficulty and discriminatory properties. It also provides a complete picture of the relation between item performance and the total score. STATITEM can display an empirical curve computed from the original scores or a smoothed version that may be used to eliminate noise caused by percentages based on a small number of subjects.

Another common discrimination index is the **point-biserial correlation** between an item and the scale total when the item is omitted. The item is removed from the total score since keeping it would produce artificially inflated correlation coefficients. A positive value suggests that those

who answered the item correctly scored relatively high on the scale total. A negative value indicates that those who answered the item correctly have obtained relatively low scores on the scale total.

The **biserial correlation coefficient** is an index derived from the point-biserial correlation. This coefficient assumes that the variable measured by a dichotomous item response is in fact a continuous variable that is normally distributed. Since the reduction of a continuous variable into a dichotomy has the effect of reducing its correlation with other variables, the computation of the biserial correlation consists of applying a correction to the point-biserial that tries to estimate what would have been the value of this correlation if the item had not been dichotomized. One drawback of this correction is that the biserial correlation is no longer bounded by the -1.0 and 1.0 limits and can take values lower than -1 or higher than 1.

INAPPROPRIATE USE OF ITEM ANALYSIS

When a scale is used to assess the knowledge of individuals on a particular subject for a purpose other than selection or discrimination, it may be inappropriate to use some of the techniques of item selection. Examples of this situation include mastery learning or any other instructional intervention that attempts to teach a new skill and for which achievement is measured using a criterion-referenced test. In these situations, a test administered prior to the instructional intervention would yield a very low percentage correct for each item and item-total correlations near zero, reflecting the fact that examinees knew nothing about the subject matter and responded almost randomly to the test. These items would indicate the examinees do need such training so that discarding the items based on their low discriminatory value or their low correlation with the total score would clearly be inappropriate. Furthermore, performing the same analysis after a successful training would produce very high percentages of success. An item successfully answered by a great majority of individuals would be considered to have little or discriminatory value and would produce either a low or an indeterminate item-total correlation. Again, discarding this item would mean the removal of a question that measures the mastery of the skill, which is exactly what we want to assess. In this situation, using the internal consistency index would also be inappropriate, since in the ideal situation where an instructional intervention brings about, for all participants, perfect mastery of a skill that was nonexistent before would mean administering a test with a Cronbach's alpha value of zero prior to the training and with a undetermined value after the training. Yet, even in this situation, item analysis may still be useful since information about response patterns may allow test developers to identify flawed items (unintentional clues, misleading formulation, etc.), and ineffective or needless instruction, or provide useful information about the kind of errors made by individuals, their misconceptions, etc.

OPTIONS

Keys location - This option allows you to specify whether the key responses are stored in the first record of the database or in a key file. If the key file option is selected, the program will look in the data file directory for a file with the same file name as the data file but with a .KEY extension. This file is a plain text file where each line contains the name of a variable followed by the value of the correct response. If a key file is not found or if a variable has no key response, the analysis will automatically stop (NOTE: Even if the responses to the item have been previously dichotomized as correct or incorrect, you still need to specify the key value for each variable by providing the value used to represent a correct response).

Exclude case with missing - This option allows you to exclude cases containing missing data (either system or user-missing values) from the analysis. If disabled, all missing data will be treated as incorrect responses.

Response total correlations - This option displays, for each item in the scale, the frequency and percentage of endorsement of each response as well as the biserial and point biserial correlation between the response and the scale total omitting that item. The item is removed from the total score since keeping it would produce artificially inflated correlation coefficients. The table also includes the value the Cronbach's alpha would take if the item was deleted from the scale as well as a discrimination index (D index) obtained by computing the difference between the percentage of correct responses in the highest and lowest groups. The percentage of respondents used for this index can be set using the **Hi-Low discrimination index** option that can take a value between 10 and 50%. To include similar information about each alternate response, enable the **Include alternate responses** option.

Item-total endorsement rate - This option allows one to obtain a matrix that displays for each value of the scale total the percentage of cases who scored positively on each item.

Item characteristic curves - This option displays curves allowing you to examine the relationship between the total score and the success level on an item, to compare the difficulty level of several items, and to estimate their discriminatory value. By default, the curves are drawn using the original scores. The **Smoothed curve** option may be used to eliminate some irregularities in the curves caused by small numbers of subjects used in the computation.

Item-total response breakdown - This option allows one to separate the whole sample into several groups of equal size and to compare the percentage of cases who scored positively on each item. The **Number of groups** specifies the number of groups to create. This number must lie between 2 and 10. For example, setting this option to 4 will create four different groups based on the obtained total score, each group representing 25% of the respondents. The program will then display a table that indicates for each item of the scale the percentage of correct responses for these four groups. A discrimination index is also computed for each item by computing the

difference in the percentage of endorsement between the groups with the highest and the lowest scores.

Scale statistics - The Scale Statistics option displays various descriptive statistics on the distribution of the scale total (mean, median, minimum and maximum, standard deviation, standard error, skewness, kurtosis, etc.). The Cronbach's alpha internal consistency statistic and the Ferguson's delta discrimination index are also computed. This option also provides summary statistics on the percentages and on the biserial and point biserial correlations between each correct response and the scale total.

Total score distribution - Activate this option to scrutinize the distribution of total scores more closely.

Save scale totals - This option allows one to save the total score computed for each subject in an ASCII data file. This data file can then be opened by SIMSTAT to perform further analysis.

Sample output of an item analysis

ITEM ANALYSIS							
	Value	Freq	Percent	Biserial	Point-Bis Correl	Alpha if Delete	25 pct Hi-Low
V1	1.0 *	259	65%	.43	.34	.674	62%
	2.0	10	3%	-.62	-.23		
	3.0	17	4%	-.15	-.07		
	4.0	114	29%	-.33	-.25		
V2	1.0	63	16%	-.18	-.12	.686	40%
	2.0	17	4%	-.41	-.18		
	3.0	3	1%	-.78	-.19		
	4.0 *	317	79%	.34	.24		
V3	1.0	15	4%	-.52	-.22	.653	72%
	2.0 *	274	69%	.63	.48		
	3.0	22	6%	-.58	-.28		
	4.0	89	22%	-.39	-.28		
V4	1.0	14	4%	-.78	-.33	.673	22%
	2.0	3	1%	-.61	-.15		
	3.0 *	378	95%	.86	.42		
	4.0	5	1%	-.72	-.21		
V5	1.0 *	364	91%	.47	.27	.683	24%
	3.0	31	8%	-.37	-.20		
	4.0	5	1%	-.69	-.20		
V6	1.0	8	2%	-.19	-.06	.697	4%
	2.0	2	1%	.03	.01		
	3.0	1	0%	.15	.02		
	4.0 *	389	97%	.12	.05		

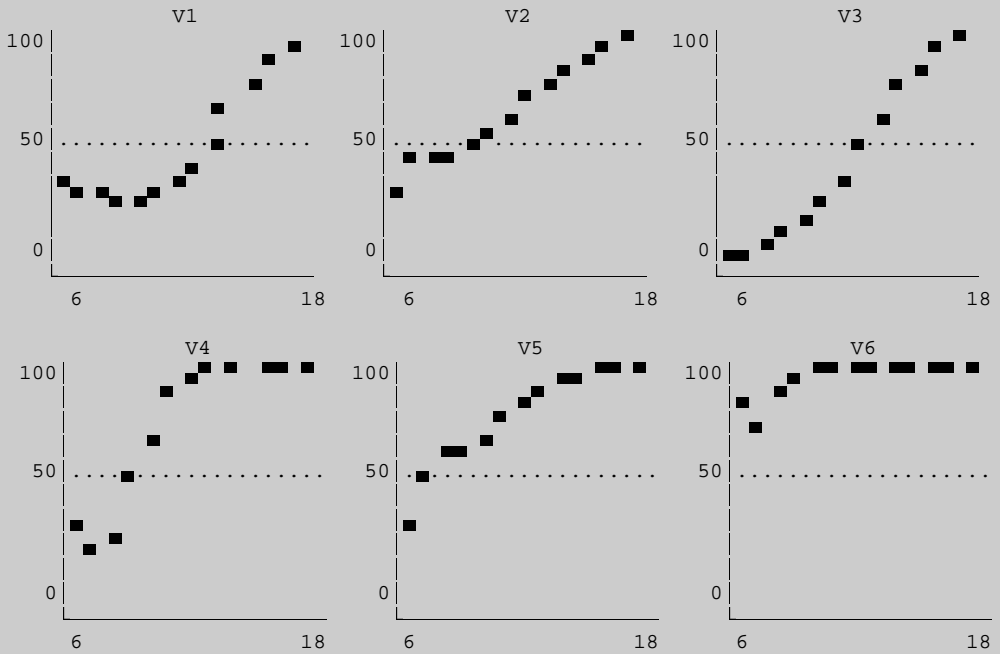
ITEM-TOTAL RESPONSE RATE

TOTAL	6	7	8	9	10	11	12	13	14	15	16
FREQUENCY	2	1	6	4	15	20	28	41	47	54	61
V1	50%	0%	33%	50%	20%	25%	39%	34%	45%	72%	80%
V2	0%	100%	50%	25%	53%	55%	54%	76%	77%	78%	85%
V3	0%	0%	0%	0%	20%	15%	29%	51%	62%	74%	85%
V4	50%	0%	17%	25%	53%	95%	86%	100%	100%	100%	100%
V5	0%	0%	83%	50%	60%	75%	86%	85%	96%	93%	97%
V6	100%	0%	83%	100%	93%	100%	100%	98%	96%	96%	95%

TOTAL	17	18
FREQUENCY	60	61

V1	85%	100%
V2	93%	100%
V3	95%	100%
V4	100%	100%
V5	98%	100%
V6	100%	100%

ITEM CHARACTERISTIC CURVES (Smoothed)



ITEM-TOTAL RESPONSE BREAKDOWN (25% of respondents in each group)

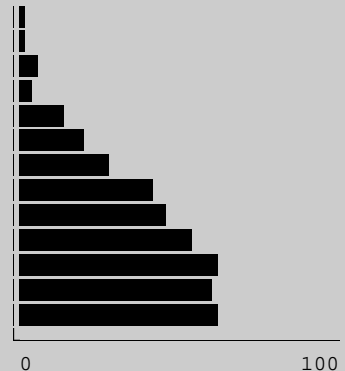
	C25	C50	C75	C100	HI-LOW
V1	32%	53%	80%	94%	62%
V2	57%	77%	86%	97%	40%
V3	26%	64%	85%	98%	72%
V4	78%	100%	100%	100%	22%
V5	75%	93%	96%	99%	24%
V6	96%	96%	96%	100%	4%

SCALE STATISTICS

	Total score	Percentage	Biserial	Item-Total
Mean	14.76	82.0	.45	.28
Median	15.00	85.8	.45	.28
Minimum	6.00	48.0	-.03	-.02
Maximum	18.00	97.3	.86	.48
Variance	6.62	206.7	.04	.01
Std. Dev.	2.57	14.4	.19	.12
Skewness	-.745			
S.E. Skewness	.122			
Kurtosis	.104			
S.E. Kurtosis	.245			
Nb of cases	400			
Nb of items	18			
Cronbach's Alpha	.694			
Ferguson's Delta	.928			

TOTAL SCORE FREQUENCY DISTRIBUTION

Value	Frequency	Percent	Cumul Frequency	Cumul Percent
6	2	.5	2	.5
7	1	.3	3	.8
8	6	1.5	9	2.3
9	4	1.0	13	3.3
10	15	3.8	28	7.0
11	20	5.0	48	12.0
12	28	7.0	76	19.0
13	41	10.3	117	29.3
14	47	11.8	164	41.0
15	54	13.5	218	54.5
16	61	15.3	279	69.8
17	60	15.0	339	84.8
18	61	15.3	400	100.0
TOTAL	400	100.0	400	100.0



Kolmogorov-Smirnov 1 Sample Test

The KOLMOGOROV-SMIRNOV one-sample test compares the distribution of each variable against a standard normal distribution or a uniform distribution. It tests whether the sample data can reasonably be thought to have come from a population having this distribution. The output displays the largest absolute positive and negative differences between the two distributions, Kolmogorov-Smirnov's Z, and the two-tailed probability test.

OPTIONS

Distribution - This option allows you to choose between distribution types: normal or uniform.

Mean - Standard Deviation - These options allow you to specify the mean and the standard deviation of the hypothetical normal distribution. If no value is given, the observed mean and standard deviation are used.

Minimum - Maximum - These options allow you to specify the minimum and the maximum values of the hypothetical uniform distribution. If no values are given, the observed minimum and maximum are used.

Sample output of a Kolmogorov-Smirnov goodness-of-fit test

KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST: AGGRESS Nb of aggressive behavior				
Type of Distribution: Normal		Mean: 27.542	Std Dev: 16.578	
Most Extreme Differences				
Absolute	Positive	Negative	K-S Z	2-tailed P
.16418	.16418	-.10451	1.261	.083

Kolmogorov-Smirnov 2 Samples Test

The KOLMOGOROV-SMIRNOV two-samples test evaluates whether a variable has the same distribution in two independent samples as defined by a grouping variable. This test is sensitive to differences in the shape, location, and scale of the two sample distributions. The output displays the count in each group, the largest absolute, positive, and negative differences between the two groups, Kolmogorov-Smirnov's Z, and the two-tailed probability for each variable.

OPTIONS

Values of x - This option requires two independent variable values. The cases that match the first value on the independent variable form one group and the cases that match the second value form the second group. The order in which values are specified determines which difference is the largest positive and which is the largest negative.

Sample output of a Kolmogorov-Smirnov two samples test

KOLMOGOROV-SMIRNOV TWO SAMPLES TEST				
AGGRESS	Nb of aggressive behavior			
With SEX	Sex of the child			
Cases				
29	SEX =	1.00		
30	SEX =	2.00		

59	Total			
Most Extreme Differences				
Absolute	Positive	Negative	K-S Z	2-tailed P
.72989	.00000	-.72989	2.803	.000

Kruskal-Wallis

The KRUSKAL-WALLIS one-way analysis of variance by ranks is a procedure for testing whether k groups have been drawn from the same population. This test is a nonparametric version of the ONEWAY analysis of variance. The output displays the number of valid cases, the mean rank of the variable in each group, chi-square and its probability with a correction for ties.

OPTION

Range of x - This option requests two values that will be treated as minimum and maximum values of the grouping (or independent) variable under consideration. Each discrete value of the independent variable that falls within this range defines a distinct group.

Sample output of a Kruskal-Wallis one-way analysis of variance by ranks

KRUSKAL-WALLIS 1-WAY ANOVA					
AGGRESS	Nb of aggressive behavior				
by SIBLING	Number of siblings				
Mean Rank	Cases				
26.45	29	SIBLING =	.00		
30.33	24	SIBLING =	1.00		
45.83	6	SIBLING =	2.00		

	59	Total			
				Corrected for Ties	
CASES	Chi-Square	Significance	Chi-Square	Significance	
59	6.3480	.0418	6.4123	.0405	

Listing cases

LIST CASES displays a listing of the values of the selected dependent and independent variables.

OPTION

List all cases - Check this box if you want to display all currently selected cases.

Number of cases - When the List All Cases option is disabled, this option allows you to specify how many cases to include in the listing.

Sample output of list cases

LISTING OF DATA (N = 10)

	SEX	AGE	SIBLING	HOURSTV	AGGRESS
1	1.00	6.00	.00	12.00	16.00
2	1.00	11.00	.00	11.50	46.00
3	1.00	9.00	.00	21.00	66.00
4	1.00	10.00	1.00	16.50	66.00
5	1.00	10.00	1.00	17.50	26.00
6	1.00	10.00	.00	15.50	26.00
7	2.00	10.00	2.00	22.50	53.00
8	1.00	11.00	1.00	15.50	46.00
9	1.00	9.00	1.00	14.00	56.00
10	1.00	9.00	.00	16.00	26.00

Logistic regression

The LOGISTIC REGRESSION command fits a multiple logistic regression model on a binary response variable with one or several explanatory variables. Output includes the likelihood ratio statistic for overall significance, parameter estimates, exponentiated parameter estimates (which are the odds ratios corresponding to a unit change in the independent variables), Wald statistics for assessing the effects of independent variables, and confidence intervals for the regression parameters (For more detailed information about logistic regression and the statistics computed, see the LOGISTIC user's manual).

NOTE: LOGISTIC v3.11Ef is a shareware program written by Dr Gerard E. Dallal. You can obtain a copy of the program on any Internet SIMTEL FTP site such as oak.oakland.edu or by writing to:

Gerard E. Dallal
54 High Plain Road
Andover, MA 01810
USA

OPTIONS

Value for success and **Value for failure** - These two options allow you to specify legal values of the binary response. By default those values are set to 0 and 1.

Constant - This option specifies whether or not to include a constant in the model.

Interaction - This option allows you to specify which interactions should be included in the model. Variables are designated by a single uppercase letter and are grouped together by an * character. Multiple interactions may be specified on the same line. For instance the following expression: A*C A*B*C designate a two-way interaction between the first and the third variables on the list, and a three way interaction between the first three variables.

Classification table - When enabled, this option constructs a classification table (observed response vs. P(observed response = 1), with probabilities, grouped in tenths, determined from the fitted regression model) to aid in assessing the adequacy of the fitted model. If 3 or more rows have positive totals, a Hosmer-Lemeshow goodness-of-fit statistic (Hosmer and Lemeshow, 1989, sec. 5.2.2) is computed along with its P-value.

Likelihood ratio - This option allows you to display the likelihood ratio statistics for the significance of each variable. They can be used as a check on the corresponding Wald statistics, which Hauck and Donner (1977) have shown to be misleading sometimes. This option was implemented as a separate command because of the work it generates: a different model must be fitted (iteratively!) to assess the effect of each variable. This option is not available when the model contains interactions.

Confidence interval - This option specifies the width of the confidence intervals for the coefficients.

Tolerance - This option specifies the convergence criterion. Iterations cease when the largest relative change in any coefficient between successive iterations is smaller than the specified tolerance. The maximum number of iterations is set at 50, and cannot be altered.

Mann-Whitney U test

The MANN-WHITNEY U test evaluates the hypothesis that two independent samples have the same distribution. The Mann-Whitney U is the nonparametric version of the t-test for independent samples. This test is performed on the dependent variable divided into two groups as defined by values of the independent (grouping) variable.

OPTIONS

Values of x - This option requires two independent variable values. The cases that match the first value on the independent variable form one group and the cases that match the second value form the second group.

Direction - This option allows you to select either a one-tailed (directional) or two-tailed (non-directional) probability test.

Sample output of a Mann-Whitney U - Wilcoxon rank sum W test analysis

MANN-WHITNEY U - WILCOXON RANK SUM W TEST				
AGGRESS	Nb of aggressive behavior			
With SEX	Sex of the child			
Mean Rank	Cases			
40.48	29	SEX = 1.00		
19.87	30	SEX = 2.00		

	59	Total		
U	W	EXACT 1-tailed P	Corrected for ties Z	1-tailed P
131.0	1174.0	.0000	-4.6325	.0000

McNemar test

The McNEMAR TEST is a procedure applied to a pair of correlated dichotomous variables to test whether there is a significant difference in proportions of subjects that change from one category to another at two different points in time, or in response to two different conditions.

A binomial test is used to compute the significance level when the total number of changes is less than 10. Otherwise, a chi-square statistic with the Yates correction for continuity is used.

OPTIONS

Values - This option allows you to specify the two values for both the independent and dependent variables. For each variable, the cases that match the first value are assigned to one condition and the cases that match the second value are assigned to the second condition. A 2 x 2 contingency table is then constructed and a significance test is computed for cases that are in different conditions.

Direction - This option specifies whether the probability of the significance test is based on a one- or two-tailed test.

Sample output of McNemar test

```
McNEMAR TEST:  VAR2 With  VAR1
```

		VAR2	
		2	1
VAR1	1	9	5
	2	3	7

Chi = .063 2 tailed probability = .8026

Median test

The MEDIAN TEST is a procedure for testing whether two or more independent groups differ in central tendencies. It tests the likelihood that those groups were drawn from populations with the same median.

The output displays the number of cases greater than, the number of cases less than, and the number of cases equal to the median for each category of the grouping variable. Also displayed are the median, chi-square, degrees of freedom and probability value.

OPTIONS

Type - The Type option allows you to choose between a design including 2 samples, or an extended version that tests for more than 2.

Values of x - You must enter two values in the Values of X option. If the design chosen is a 2 samples test, then two groups are formed using the two values. If the extended test is chosen, then every value in the range defined by the two values forms a group. A test is then performed on all groups.

Sample output of a median test

MEDIAN TEST				
AGGRESS	Nb of aggressive behavior			
With SIBLING	Number of siblings			
	Gt Median	Le Median		
SIBLING = .00	7	22		
SIBLING = 1.00	11	13		
SIBLING = 2.00	4	2		
Cases	Median	Chi-square	D.F.	Significance
59	26.000	5.1086	2	.0777

Moses test of extreme reactions

The MOSES TEST of extreme reactions tests whether the range of an ordinal variable is the same in a control group as in a comparison group, as defined by a grouping variable. The output includes counts for both groups, number of outliers removed, the span of the control group before and after outliers are removed, and one-tailed probability of the span with and without outliers.

OPTIONS

Values of x - This option requires two independent variable values. The first value of the independent variable defines the control group, and the second value defines the comparison group.

Outliers to remove - This option allows you to determine the percentage of extremes cases to be excluded from the analysis. If this field is left blank, 5% of the cases are trimmed from each end of the range of the control group to remove outliers.

Sample output of a Moses test of extreme reactions analysis

```

MOSES TEST OF EXTREME REACTIONS

      HOURSTV   Hours per week spent watching TV
With SEX      Sex of the child

              Cases
      (CONTROL)    29  SEX = 1.00
      (EXPERIMENTAL) 30  SEX = 2.00
              -----
              59  Total

1-tailed P      Span of Control Group
      .0293      52  Observed
      .1633      50  After removing 1 Outlier(s) from each end

```

Multiple regression

Multiple regression analysis is a statistical technique that allows you to assess the relationship between one dependent (Y) variable and several independent variables (also called predictors). For example, you may want to predict the aggressiveness of children from several independent variables such as gender, age, number of siblings and the time spent watching TV. The technique can be used to analyze how various combinations of independent variables correlate with one another as well as with the dependent variable.

Multiple regression is an extension of bivariate regression. The result of regression is an equation that represents the prediction of the dependent variable from several independent variables. The regression equation takes the following form:

$$Y' = A + B_1X_1 + B_2X_2 + \dots + B_kX_k$$

where Y' is the predicted value of the dependent variable, A is the intercept (the value of Y when all values of the independent variables are zero), X represents the observed value of the independent variables (or predictors) and B is the coefficient assigned to each of the independent variables. The goal of the regression technique is to find the values of all B that produce prediction scores that most closely fit the actual values of Y.

SIMSTAT provides three broad classes of multiple regression: standard regression, hierarchical regression and statistical (stepwise) regression. Each class differs in the way the independent variables are selected to be included in the equation.

Standard Multiple Regression

In standard regression, all independent variables are entered into the regression equation at the same time. In this type of regression, some independent variables may be entered in the model even if their simple correlation with the dependent variable is negligible. It is also possible that some variables with initial high correlation with the dependent variable are entered, but since they are also highly correlated with other predictors, their unique contribution to the prediction of the dependent variable becomes very low.

Hierarchical Multiple Regression

The hierarchical multiple regression model allows you to specify the order in which the independent variables will be entered in the equation. The variables can be entered one variable at a time or in sets of variables. At each step, information can be generated about the degree to which this new variable or set of variables adds to the prediction of the dependent variable. The order of entry of variables is chosen according to theoretical or logical considerations such as the theoretical importance, causal precedence or elimination of 'nuisance'.

Statistical Regression

In the statistical regression model, the independent variables are entered in the equation according solely to some statistical criteria obtained from the current sample. SIMSTAT provides three versions of statistical regression: forward selection, backward deletion, and stepwise selection. When forward selection is used, the variable that shows the largest correlation (positive or negative) with the dependent variable is entered in the equation provided that it meets a specified statistical criterion of significance. Then, a comparison is made between all the remaining variables to select the one that contributes the most to a significant increase in prediction. Forward selection continues until there are no other variables that meet the entry criterion. In backward deletion, all variables are entered initially. Then, variables that do not meet the statistical criterion are sequentially removed. Stepwise selection is a combination of the forward and backward procedures. With this procedure, a variable is selected to be entered in the equation in the same manner as with forward selection, but after each entry of a new variable, each variable already in the equations is examined so that if it no longer contributes significantly to the regression, it is removed. The criteria used to enter or remove a variable from the equation can be specified by the user.

Much controversy surrounds the use of this type of multiple regression. One of the criticisms of this technique is that it capitalizes on chance and thus offers a solution that often overfits the sample data and cannot be generalized to the population or even be replicated on another sample. In a sense, the solution obtained from the sample may be very unstable.

The reason that those procedures have been included here is that there remain some rare situations in which these techniques are needed, such as when one wants to select a limited number of variables among a set of good predictors (mainly for practical reasons). In the author's opinion, hierarchical regression is probably the most highly recommended procedure. However, this procedure requires clarification of the logic and theory behind the data collection.

OPTIONS

ANALYSIS PAGE

Method - The Method list box option allows you choose among 5 different strategies of multiple regression: Hierarchical entry, forward selection, backward deletion, stepwise selection, and standard regression.

Show steps - This option requests the printing of various statistics at each step of the regression. The information output at each step can include an ANOVA table, a test for change, and various statistics for the variables in the equation and for those not yet in the equation.

ANOVA table - The ANOVA Table option allows printing of an ANOVA table that includes regression analysis and residual sum of squares, mean square, F and probability value associated with F.

Test of change - This option allows printing an ANOVA table that tests whether the new variable(s) in the model significantly increase R square above the R squared predicted with the variables already in the equation. The table includes the resultant R squared, the R squared change, the sum of squares, the F ratio and its probability value.

Variables in the equation - This option displays various statistics computed on the variables in the equation including each unstandardized regression coefficient (or B_1), its standard error and confidence limits, the standardized coefficient (or beta), whole, partial and semi-partial correlation coefficients, tolerance level, F ratio and its probability.

Variables not in the equation - This option displays the tolerance level, the F ratio and its probability for each variable not yet in the equation.

History - When the History option is activated a summary report of the changes in R squared at each step of the regression is printed.

Summary ANOVA - This option allows the printing of a detailed ANOVA table that includes the mean square, the sum of squares, the F ratio and its probability for each variable or set of variables in the equation.

Confidence interval - This option allows you to set the width of the confidence interval for the unstandardized regression coefficients. This interval is expressed as a percentage and must lie between 1% and 99%.

DIAGNOSIS PAGE

Residual caseplot - This option allows the output of a casewise plot of standardized residuals, including the predicted, obtained, and residual values for all cases. This option is useful for identifying outliers (i.e., cases that are not well represented by the regression model). The **Outliers** option allows you to restrict the residual caseplot to those cases for which the absolute standardized value is greater than or equal to the specified value. The Outliers value can be set to between 0 and 4 standard deviations.

Durbin-Watson statistic - This option tests for the presence of autocorrelation or serial correlation in the residuals. The larger the autocorrelation, the less reliable the results of the analysis.

Residual scatterplot - This procedure produces a bivariate scatterplot. The predicted value is plotted along the horizontal axis, and the standardized residuals along the vertical axis.

Residual normal plot - This option displays a normal probability plot of residual values that allows you to evaluate whether those residuals are normally distributed. If the residuals follow a normal distribution, the data points will fall approximately along a straight line going from the lower left corner of the graph to the upper right corner.

Save residuals - This option instructs SIMSTAT to save the predicted and residual values as new variables in the current data file. Creating those variables allows you to perform further analyses on them such as displaying scatterplots between each independent variable and the predicted and residual values. The new variables are named PREDxxx and RESIDxxx where xxx stands for a serial number between 001 and 999 that corresponds to the number of regression analyses performed during a single command. This number is automatically reset to 001 after each command. To prevent overwriting those variables, you must rename them using the DEFINE VARIABLE command.

CRITERIA PAGE

P to enter - P to remove - The P to Enter and P to Remove fields allow you to set criteria to be used in statistical regression techniques (i.e. stepwise, forward and backward methods). A variable will be entered in the model if its probability is less than the P to Enter criteria, while a variable in the equation that has a probability above the P to Remove criteria will be removed from the model. The P to Remove should always be greater than the P to Enter value.

Tolerance - The Tolerance criterion is used to prevent the inclusion of a variable that would produce multicollinearity in the equation. This measure is the proportion of variance of the variable not explained by the other independent variables already in the equation (or $1 - R^2$). In order to be entered or to remain in the equation, a variable must pass this tolerance test. To disable this function, set the criterion value to zero.

ORDER PAGE

When a hierarchical multiple regression method is chosen, the dialog box displays an Order tab with a grid that lets you specify the order of entry of each independent variable in the regression model. The order number can lie between 1 and 20. All variables with the same number are entered at the same time, while variables that have lower numbers are entered before those that have higher numbers. Variables with identical order numbers will be entered simultaneously. You can enter the order by using the keyboard or the spin button located at the right edge of the grid. At each step, information can be generated about the degree to which a new variable or set of variables explains variance in the dependent variable.

Sample output of a hierarchical multiple regression

```
MULTIPLE REGRESSION
Dependent Variable:  AGGRESS  Nb of aggressive behavior
Method: Hierarchical entry
***** STEP 1 *****
Variable(s) entered on Step 1      SEX  Sex of the child
```

	AGE	Age of the child					
	SIBLING	Number of siblings					
Significance test for change							
Source	D.F.	Sum of Squares	Rsq Chg	F Ratio	F Prob.		
New variable(s)	3	6885.2089	.4319	13.940	.0000		
Regression	3	6885.2089					
Residual	55	9055.4351					
Multiple Regression							
Multiple R	= .6572	sig. of R = .0000					
Multiple R Square	= .4319						
Adjusted R square	= .4009						
Analysis of Variance							
Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.		
Regression	3	6885.2089	2295.0696	13.940	.0000		
Residual	55	9055.4351	164.6443				
Equation: AGGRESS = 21.9872 + (-16.5393 * SEX) + (2.8501 * AGE) + (7.0595 * SIBLING)							
Variables in the equation							
Variable	B	SE B	80% confidence interval		Tolerance		
Intercept	21.9872						
SEX	-16.5393	3.3674	-20.9045 to	-12.1741	.9847		
AGE	2.8501	1.1501	1.3593 to	4.3410	.9829		
SIBLING	7.0595	2.5298	3.7802 to	10.3389	.9882		
Variable	Beta	SE Beta	Correl	S-Part	Partial	F	Sig F
SEX	-.5030	.1024	-.5471	.4992	-.5522	24.124	.0000
AGE	.2540	.1025	.2813	.2519	.3169	6.141	.0163
SIBLING	.2853	.1022	.2988	.2836	.3522	7.787	.0072
Variables not in the equation							
Variable	F	Sig F					
HOURSTV	3.650	.0613					

```

***** STEP 2 *****
Variable(s) entered on Step 2  HOURSTV  Hours per week spent watching TV
Significance test for change

Source          D.F.          Sum of
                Squares          Rsq Chg          F          F
                Ratio          Prob.

New variable(s)  1            573.2781          .0360          3.650  .0614

Regression      4            7458.4870
Residual        54           8482.1570

Multiple Regression

Multiple R      =  .6840          sig. of R =  .0000
Multiple R Square =  .4679
Adjusted R square =  .4285

Analysis of Variance

Source          D.F.          Sum of
                Squares          Mean
                Squares          F          F
                Ratio          Prob.

Regression      4            7458.4870          1864.6218          11.871  .0000
Residual        54           8482.1570          157.0770

Equation: AGGRESS = 11.6725 + (-16.5099 * SEX) + (2.6390 * AGE)
              + (5.7389 * SIBLING) + (.8717 * HOURSTV)

Variables in the equation

Variable          B          SE B          80% confidence interval  Tolerance

Intercept        11.6725
SEX              -16.5099          3.2892          -20.7737 to  -12.2462          .9846
AGE              2.6390          1.1288          1.1758 to  4.1022          .9735
SIBLING          5.7389          2.5658          2.4127 to  9.0650          .9165
HOURSTV          .8717          .4563          .2802 to  1.4632          .9217

Variable          Beta  SE Beta  Correl  S-Part  Partial          F          Sig F

SEX              -.5021  .1000  -.5471  .4983  -.5640  25.195  .0000
AGE              .2352  .1006  .2813  .2321  .3032  5.466  .0231
SIBLING          .2319  .1037  .2988  .2220  .2912  5.003  .0295
HOURSTV          .1975  .1034  .2921  .1896  .2516  3.650  .0614
    
```

Summary Analysis of Variance

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
BLOCK #1	3	6098.7331	2032.9110	12.942	.0000
HOURSTV	1	573.2781	573.2781	3.650	.0614
Explained	4	7458.4870	1864.6218	11.871	.0000
Residual	54	8482.1570	157.0770		
Total	58	15940.6441	274.8387		

Summary tests for change

Step	Variable(s)	R Square	Rsq Chg	F Chg	D.F.	Prob.
1	+ SEX,AGE,SIBLING	.4319	.4319	13.940	3,55	.0000
2	+ HOURSTV	.4679	.0360	3.650	1,54	.0614

Standardized residual caseplot

Case #	Predicted	Obtained	Residual	-3.0	0.0	3.0
				:	:	:
				:	:	:
Case #	Predicted	Obtained	Residual	-3.0	0.0	3.0

Durbin-Watson test = 1.5783

Multiple responses analyses

The MULTIPLE RESPONSE procedure allows you to obtain descriptive analyses and crosstabulation analyses on variables which can legitimately have more than one response. These multiple responses are stored in as many variables as necessary.

Suppose you conduct a survey on the brand of cereal that peoples ate for breakfast in the previous week, allowing respondents to give up to 3 brands of cereal. One way to code this information in a data file would be to create as many variables as there are brands of cereal and record for each of those variables whether the respondent ate that particular brand or not. Another way would be to create 3 variables corresponding to the maximum number of responses and to enter a code corresponding to each brand of cereal. One problem with this type of coding is that in order to obtain a simple frequency analysis you would have to perform 3 different frequency analyses and compute the total frequency by manually aggregating those 3 tables. The difficulties increase if you choose to perform a crosstabulation of this variable with another one.

The MULTIPLE RESPONSES procedure allows you to easily obtain frequency and crosstabulation analyses by treating responses stored in several variables as if they were stored in a single variable. The dependent (X) and independent variable (Y) designations (see Choose X & Y command) are used to gather all the variables together.

OPTIONS

Independent/dependent variables as multiple responses - These two options allow you to specify which type of variable is to be treated as multiple responses.

Pressing the OK button accesses a second dialog box that allows you to alter any options normally available in the standard FREQUENCY, CROSSTAB or INTER-RATERS analysis.

Sample output of Multiple responses frequencies

MULTIPLE RESPONSE FREQUENCIES: NBREST1		Nb of meals in restaurants this week		Nb of meals in restaurants last week		
	Value	Frequency	Pct of Responses	Pct of Cases	Valid Pct of Responses	Valid Pct of Cases
None	0	29	24.6	49.2	24.6	49.2
	1	53	44.9	89.8	44.9	89.8
	2	36	30.5	61.0	30.5	61.0
	TOTAL	118	100.0	200.0	100.0	200.0
Mean	1.059	Variance	.552	Kurtosis	-1.169	
Median	1.000	Std Dev	.743	S.E. Kurt.	.451	
Mode	1.000	Std Err	.068	Skewness	-.096	
Minimum	.000	Sum	125.000	S.E. Skew.	.225	
Maximum	2.000	Range	2.000	Valid	118.000	
95% Confidence Interval for the mean = [.9253 to 1.1934]						

Sample output of a multiple responses crosstab analysis

MULTIPLE RESPONSE CROSSTAB: DAY	Day of the week
---------------------------------	-----------------

		by NBREST1		Nb of meals in restaurant this week		
		NBREST2		Nb of meals in restaurant last week		
NBREST1	NBREST2 -->					
Subject	Count Tot Pct					Total
		.00	1.00	2.00		
DAYS	1	1	4	3		8
		.8	3.4	2.5		6.8
		1.7	6.8	5.1		
	2		3	1		4
		.0	2.5	.8		3.4
		.0	5.1	1.7		
	3	3	2	7		12
		2.5	1.7	5.9		10.2
		5.1	3.4	11.9		
	4	10	11	5		26
		8.5	9.3	4.2		22.0
		16.9	18.6	8.5		
	5	7	11	10		28
		5.9	9.3	8.5		23.7
		11.9	18.6	16.9		
	6	8	22	10		40
		6.8	18.6	8.5		33.9
		13.6	37.3	16.9		
Column		29	53	36		118
Total		24.6	44.9	30.5		100.0

Nonparametric matrix

The NPAR MATRIX displays a matrix for various measures of association and concordance between two variables.

OPTIONS

Estimator - The ESTIMATOR option allows you to choose from a drop down list which measures of association are to be computed and displayed in an association matrix. SIMSTAT currently supports the following measures:

- Kendall's tau-a
- Kendall's tau-b
- Kendall-Stuart's tau-c
- Somers' d (symmetric)
- Somers' d_{xy} (asymmetric)
- Somers' d_{yx} (asymmetric)
- Goodman-Kruskal's gamma
- Spearman's r_s
- Pearson's r

Type of matrix - When set to X vs Y, the association matrix displays relationships between all variables assigned as independent against all variables assigned as dependent. The square matrix option produces a matrix displaying relationships between all selected variables, without taking into account whether they were selected as dependent or independent.

Display probability value - When this option is enabled, SIMSTAT outputs an association matrix with counts and exact probabilities for each coefficient. When disabled, asterisks are used to show the probability level attained. The standard matrix includes the chosen coefficient with up to 3 asterisks (*) corresponding to the significance level attained.

Significance test - This option allows you to specify whether the probability of the coefficients is based on a one-tailed (directional) or two-tailed (non-directional) test.

Missing values - This option allows you to specify whether you want to exclude cases with missing values by either PAIRWISE or LISTWISE deletion. If you select pairwise deletion, a case is excluded if it has a missing value on either of the two variables used to compute a given coefficient. However, this case can be included in the computation of the other coefficients. Listwise deletion excludes cases containing missing data from the computation of all the measures included in the matrix.

Sample output of a nonparametric association matrix (tau-c)

ASSOCIATION MATRIX: Kendall-Stuart's Tau-c					
	SEX	AGE	SIBLING	HOURSTV	AGGRESS
SEX	.9997 (59) P= .000	-.1023 (59) P= .243	-.0781 (59) P= .284	-.0046 (59) P= .488	-.6986 (59) P= .000
AGE	-.1023 (59) P= .243	.9170 (59) P= .000	.0095 (59) P= .467	.0889 (59) P= .191	.1930 (59) P= .028
SIBLING	-.0781 (59) P= .284	.0095 (59) P= .467	.8739 (59) P= .000	.2646 (59) P= .013	.2439 (59) P= .019
HOURSTV	-.0046 (59) P= .488	.0889 (59) P= .191	.2646 (59) P= .013	.9809 (59) P= .000	.1518 (59) P= .049
AGGRESS	-.6986 (59) P= .000	.1930 (59) P= .028	.2439 (59) P= .019	.1518 (59) P= .049	.9604 (59) P= .000

(Coefficient / (Case) / Probability 1-tailed)

One sample chi-square test

The ONE SAMPLE CHI-SQUARE TEST allows you to assess whether there is a difference between the observed number of cases in various categories and the expected frequencies in those same categories. The dialog box allows you to restrict the test to specific values and to specify the expected frequencies.

OPTIONS

Values - This option allows you to specify the values that define the various categories. For example, if you enter the following string:

```
1 1.5 2 3
```

the analysis will be restricted to the cases with values equal to one of these four categories. This option can also be used to include categories for which there are no cases. If you leave this field blank, all distinct values encountered in the variable will form a separate category.

Frequency - This option allows the specification of the distribution against which the sample will be tested. All values (expected frequencies, percentages or proportions) are transformed into relative frequencies. The number of values provided in this field must match the number of categories in the values field or in the data file. If this option is left blank, equal frequencies are assumed for all categories.

Sample output of one sample chi-square analysis

CHI-SQUARE TEST: SIBLING Number of siblings		
Value	Frequency	Expected
.00	29	19.67
1.00	24	19.67
2.00	6	19.67
Chi-Square =	14.881	D.F. = 2 P = .0000

Oneway ANOVA

The ONEWAY ANOVA procedure performs a one-way analysis of variance for a dependent variable on groups defined by a categorical independent variable. It allows testing whether the means of the groups (2 or more) are significantly different from each other.

ONEWAY produces a table including: between- and within-groups sums of squares, mean squares, degrees of freedom, F-ratio and its associated probability. You can also obtain for each group, descriptive statistics including count, mean, standard deviation, standard error and a user-specified confidence interval for the mean.

OPTIONS

Range of x - This option requests two values that will be treated as the minimum and maximum values of the grouping (or independent) variable under consideration. Each discrete value of the independent variable that falls within this range defines a distinct group. If those fields are left blank, all values of the independent variable will be included. An ANOVA test is then performed on these groups.

Descriptives - The Descriptives option displays the count, mean, standard deviation, standard error, and a user-specified confidence interval around the mean of the dependent variable for each group.

Post hoc tests - This option allows you to perform a post hoc multiple comparison between all group means. The program gives a choice between four different methods each one using different criterion for computing the significance level or constructing confidence intervals of the difference between two means. The *Least-significant difference (LSD)* test computes a confidence interval and a standard Student's t test between all group means. It is the most powerful a posteriori contrast test, but as the number of pairwise comparisons increases, so does the probability that at least one of the confidence intervals or significance tests is in error. For this reason, this test is usually recommended only when it is applied to comparisons that are planned before observing the data. The *Newman-Keuls* test applies different criteria depending on the number of steps between the two group means. The further apart those two group means are from each other, the larger the difference between those two groups must be in order to be significant. This procedure should be used only when the group sizes are equal. The *Tukey's honestly significant difference (HSD)* test uses a single criterion for all comparisons regardless of the distance between the group means. This test keeps the experiment-wise error rate equal to α . The *Scheffé's* test is used to obtain simultaneous confidence intervals for differences between all pairs of means while keeping the overall error rate to α . It is a more conservative test and will lead to fewer significant

differences. This test can be used with equal or unequal group sizes, and with planned or unplanned comparisons.

Confidence interval - This option allows you to set the confidence interval width for the means, and for the pairwise comparisons of those means (post hoc tests). This interval width is expressed as a percentage and must lie between 1% and 99%. This value is also used when displaying bar charts with a confidence interval, or an error bar graph representing a confidence interval around the mean.

Mean/Error bar graph - This option displays a mean bar and/or an error bar representing the variability of the mean, or the values in each group.

Type of error - This option allows you to select whether the error bar will represent the standard deviation, the standard error or a user-defined confidence interval. The width of this interval is set by the interval option.

With bar chart - This option allows you to draw solid bars where each bar represents the mean of a separate group.

Upper error bar only - When the bar chart option is selected and an error bar is requested, this option allows you to specify whether the error bars displayed with the mean bars should be displayed above and below the mean, or only above it.

Link means - When chosen, this option connects the means with a line.

Deviation chart - This option displays a bar chart where each bar represents the deviation of the group mean from the grand mean.

Sample output of a one-way analysis of variance

ONEWAY						
AGGRESS by SIBLING		Nb of aggressive behavior Number of siblings				
Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.	
Between Groups	2	1902.18	951.09	3.7939	.0285	
Within Groups	56	14038.46	250.69			
Total	58	15940.64				
Proportion of Variance Explained (R-Square) = .1193						
Group	Count	Mean	Std Dev	Std Err	90 Pct C.I. for Mean	
SIBLING = 0.00	29	25.28	14.83	2.75	20.59 To	29.96
SIBLING = 1.00	24	28.42	16.90	3.45	22.50 To	34.33
SIBLING = 2.00	6	44.83	16.18	6.61	31.52 To	58.14
Total	59	28.54	16.58	2.16	24.93 To	32.15
Tukey's HSD Multiple Comparisons						
SIBLING		Difference	90% Pct conf interval		Sig.	
0.00	1.00	3.1408	-6.0130 To	12.2946	.7535	
0.00	2.00	19.5575	4.6800 To	34.4349	.0214	
1.00	2.00	16.4167	1.2759 To	31.5575	.0683	

Regression analysis

REGRESSION produces a simple regression analysis for each pair of dependent-independent variables. The output includes the Pearson product-moment correlations, the intercept and slope of the regression line, and an ANOVA table for the equation. The dialog box allows you to obtain bivariate scatterplots, to select one-tailed (directional) or two-tailed (non-directional) tests of probabilities, and to request standardized residuals plots.

OPTIONS

Type of analysis - This option allows you to choose between 8 types of nonlinear regression. The following types of regression will allow the assessment of different degrees of curvilinearity among the dependent-independent relations or to obtain an equation that expresses various forms of bivariate relations. The following table presents the types of regression and their corresponding equations:

TYPE	REGRESSION EQUATION
Linear	$Y = a + b_1x$
Quadratic	$Y = a + b_1x + b_2x^2$
Cubic	$Y = a + b_1x + b_2x^2 + b_3x^3$
4th degree polynomial	$Y = a + b_1x + b_2x^2 + b_3x^3 + b_4x^4$
5th degree polynomial	$Y = a + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5$
Inverse	$Y = a + b_1 / x$
Logarithmic	$Y = a + b_1 * \ln(x)$
Exponential	$Y = ab^x$

Confidence interval - This option allows you to set the confidence interval for beta weight estimates. This interval width is expressed as a percentage and must lie between 1% and 99%.

Significance test - This option specifies whether the probability of the correlations probabilities is based on a one-tailed (directional) or two-tailed (non-directional) test.

Scatterplot - Option Scatterplot produces a bivariate scatterplot with the independent variable plotted along the horizontal axis, and the dependent variable along the vertical axis.

Residual caseplot - When checked, this option produces a casewise plot of standardized residuals, including the predicted, obtained, and residual values. The **Outliers** value

allows you to restrict the residual caseplot to those cases for which the absolute standardized value is greater or equal to the specified value. The Outliers value can be set to between 0 and 4 standard deviations.

Durbin-Watson statistic - This option tests for the presence of autocorrelation or serial correlation in the residuals. The larger the autocorrelation, the less reliable the results of the analysis.

Residual scatterplot - This procedure produces a bivariate scatterplot where the predicted value is plotted along the horizontal axis, and the standardized residuals along the vertical axis. The graph can be printed in text mode, graphics mode, or both.

Residual normal plot - This option displays a normal probability plot of the residuals that allows you to evaluate whether those data are normally distributed. If the residuals follow a normal distribution, the data points will fall approximately along a straight line going from the lower left corner of the graph to the upper right corner.

Save residuals - This option instructs SIMSTAT to save the predicted and residual values as new variables in the current data file. Creating those variables allows you to perform further analyses such as displaying scatterplots between each independent variable and the predicted and residual values. The new variables are named PREDxxx and RESIDxxx where xxx stands for a serial number between 001 and 999 that corresponds to the number of regression analyses performed during a single command. This number is automatically reset to 001 after each command. To prevent overwriting those variables, you must rename them using the DEFINE VARIABLE command.

Sample output of a regression analysis

```

REGRESSION

      AGGRESS  Nb of aggressive behavior
With HOURSTV  Hours per week spent watching TV

Regression

      R = .3002      R Square = .0901      sig. of R = .0711
Analysis of Variance

      Source          D.F.          Sum of          Mean          F          F
                        Squares        Squares        Ratio        Prob.

      Regression      2          1436.5105      718.2552      2.773      .0711
      Residual        56         14504.1336      259.0024

Equation: AGGRESS = -4.9559 + (3.2271 * HOURSTV) + (-.06251 * HOURSTV^2)

Variable          B          SE B          80% confidence interval          F          Sig F
Intercept         -4.9559
Degree 1          3.2271          3.6046          -1.4455 to          7.8998          .802          .3745
Degree 2         -.06251          .1148          -.2114 to          .08634          .296          .5883

Standardized residual plot

Case #  Predicted  Obtained  Residual  -3.0  0.0  3.0
      4    30.2738   66.0000   35.7262  .      .      *  .
     11    31.3756   66.0000   34.6244  .      .      *  .
Case #  Predicted  Obtained  Residual  -3.0  0.0  3.0

Durbin-Watson test = 1.1178
    
```

Reliability analysis

Multiple-item additive scales are often used to measure various characteristics of a subject. One desirable feature of such scales is their reliability. The reliability of a scale refers to the consistency of the scores obtained when the scale is administered to the same group of subjects at different occasions, under different conditions, or using different sets of equivalent items that are supposed to measure the same underlying variable. By using this type of procedure, the difference or fluctuation observed between the various administrations of the test, provided that it cannot be attributed to real changes in the subject, can be used to estimate the proportion of the total variance of the test score that can be attributed to error of measurement. Various methods are used to estimate the reliability of a test.

The TEST-RETEST reliability method is obtained by administering the same test to the same subjects on two different occasions (usually no more than 2 months apart). The correlation between the scores obtained by the same subjects on the two administrations of the test is then computed to obtain a measure of temporal stability (or test-retest reliability). Provided that the length of the interval between the two administrations is short enough to preclude any real change in the variable being measured, any fluctuation between the two scores is attributed to random error of measurement. The higher the reliability, the less susceptible the scores are to the random daily changes in the condition of the subject or the testing environment. A test-retest reliability method can also be applied using alternative forms of the test. The reliability coefficient obtained with such a method measures both the temporal stability of the test and the consistency of response to different item samples.

Another method to measure the reliability of a test, known as SPLIT-HALF reliability, requires only a single administration of a test. In this procedure, the items in the test are divided into two halves comparable in terms of content, difficulty, etc.. The most common splitting procedure consists of comparing the scores on the odd and even items of the test. The correlation between the two scores obtained on the same subjects is then computed. The split-half reliability coefficient provides a measure of the internal consistency of the scale.

The INTERNAL CONSISTENCY method also requires only a single administration of a test. It is based on the correlations obtained between all items of the scale. It provides an evaluation of the homogeneity of the scale (also called internal consistency). The internal consistency coefficient has been shown to be mathematically equivalent to the mean of all split-half coefficients obtained from different splits of a scale.

The RELIABILITY command also provides a method of assessing the quality of multiple-item additive scales through the computation of split-half reliability and internal consistency statistics. Test-retest reliability can be assessed using the REGRESSION or CORRELATION procedures. Each selected variable is considered as a single item of the scale. The Xs and Ys are used in the split half method to specify how the various items should be divided.

The dialog box provides a choice of various item statistics (e.g., mean, minimum, maximum, standard deviation), inter-item variance-covariance and correlation matrices, total scale and item-total statistics. The maximum number of items is limited to 90 items per scale.

OPTIONS

Item statistics - This option allows the display of summary statistics on each item in the scale including the item mean, minimum, maximum, standard deviation and mean correlation with other items. It also provides various statistics for the whole scale such as the item means, variances, covariances and correlations.

Inter-item correlations - This option displays an inter-item correlation matrix that allows the identification of items that have small correlations with the other items.

Variance/covariance - This option computes an inter-item variance-covariance matrix.

Item-total statistics - This option displays various statistics that allow the evaluation of the effect of each item on the reliability of the scale. The output consists of the average score of the scale and its variance when the item is excluded, the correlation of the scores on this item with the sum of the scores of the remaining items, and the Cronbach's alpha that would result from the exclusion of this item.

Split-half reliability - This option allows one to assess the correlation between two parts of the scale. When selecting the variables, the Xs and Ys can be used to split the various items into two halves. The output contains summary statistics for both scales (mean, variance and standard deviation), the correlation between those scales, the Spearman-Brown coefficient, and the Guttman split-half coefficient which does not assume that the two parts have the same variance.

Internal consistency - This option allows the computation of Cronbach's alpha internal consistency statistics. It has been shown to be equivalent to the average of all possible split-half coefficients. The output also includes the mean inter-item correlation, and the standardized item Alpha (the alpha that would have been obtained if all the items had been standardized).

Sample output of reliability analysis

```
RELIABILITY ANALYSIS

DEP1      Depression scale item #1
DEP2      Depression scale item #2
DEP3      Depression scale item #3
DEP4      Depression scale item #4
DEP5      Depression scale item #5
DEP6      Depression scale item #6
DEP7      Depression scale item #7
DEP8      Depression scale item #8
```

Items statistics

	MEAN	STD DEV	MEAN CORR	N
DEP1	1.1167	.3301	.2575	1045
DEP2	1.5780	.5169	.2464	1045
DEP3	1.2325	.4382	.2504	1045
DEP4	1.1646	.3911	.2295	1045
DEP5	1.1110	.3378	.2220	1045
DEP6	1.5426	.5916	.2017	1045
DEP7	1.1799	.4444	.3119	1045
DEP8	1.5254	.5811	.2563	1045
Total	10.4507	2.0758	.3875	1045

	MEAN	MINIMUM	MAXIMUM	VARIANCE
Item mean	1.3063	1.1110	1.5780	.0419
Item variance	.2150	.1090	.3500	.0089
Inter-item corr.	.2269	.1586	.3399	.0029
Inter-item covar.	.0462	.0258	.0830	.0003

Correlation matrix

	DEP1	DEP2	DEP3	DEP4	DEP5	DEP6
DEP1	1.0000	.1992	.1830	.2813	.2531	.1805
DEP2	.1992	1.0000	.2561	.1828	.1698	.1920
DEP3	.1830	.2561	1.0000	.1956	.1749	.1779
DEP4	.2813	.1828	.1956	1.0000	.1951	.1643
DEP5	.2531	.1698	.1749	.1951	1.0000	.1729
DEP6	.1805	.1920	.1779	.1643	.1729	1.0000
DEP7	.3399	.3017	.3358	.3089	.2369	.2077
DEP8	.2242	.2764	.2796	.1586	.2296	.2037
	DEP7	DEP8				
DEP1	.3399	.2242				
DEP2	.3017	.2764				
DEP3	.3358	.2796				
DEP4	.3089	.1586				
DEP5	.2369	.2296				
DEP6	.2077	.2037				
DEP7	1.0000	.2716				
DEP8	.2716	1.0000				

Variance-covariance matrix

	DEP1	DEP2	DEP3	DEP4	DEP5	DEP6
DEP1	.1090	.0340	.0265	.0363	.0282	.0353
DEP2	.0340	.2671	.0580	.0370	.0296	.0587
DEP3	.0265	.0580	.1920	.0335	.0259	.0461
DEP4	.0363	.0370	.0335	.1530	.0258	.0380
DEP5	.0282	.0296	.0259	.0258	.1141	.0345
DEP6	.0353	.0587	.0461	.0380	.0345	.3500
DEP7	.0499	.0693	.0654	.0537	.0356	.0546
DEP8	.0430	.0830	.0712	.0361	.0451	.0700

Item-total statistics

	MEAN IF DELETED	VAR. IF DELETED	ITEM-TOTAL CORREL.	ALPHA IF DELETED
DEP1	9.3340	3.6939	.3990	.6577
DEP2	8.8727	3.3028	.3935	.6533
DEP3	9.2182	3.4638	.4005	.6519
DEP4	9.2861	3.6355	.3491	.6637
DEP5	9.3397	3.7456	.3437	.6664
DEP6	8.9081	3.2847	.3146	.6799
DEP7	9.2708	3.3145	.4926	.6306
DEP8	8.9254	3.1343	.4068	.6520

Split-half reliability

	MEAN	VARIANCE	STD. DEV.	NB OF ITEMS
Part 1	5.0919	1.0824	1.1716	4
Part 2	5.0919	1.2725	1.6192	4
Scale	10.4507	2.0758	4.3091	8
Split-half correlation =	.5512			
Spearman-Brown =	.7106			
Guttman split-half =	.7047			

Reliability statistics

Mean inter-item correlation =	.2269
Cronbach's Alpha =	.6866
Standardized Item Alpha =	.7013

Runs test

The RUNS TEST is a procedure that can be used to test whether the ordered sequence in which observations were obtained is random. In order to be carried out this test requires that all values be dichotomized. The dialog box allows you to separate observations in two distinct groups using the mean, the median or a user-specified value as a cutoff point.

OPTIONS

Cutoff point - This option allows you to specify how the values of a variable will be dichotomized. The cutoff point used can be either the mean, the median or a user-specified value. All observations falling below the cutoff point form one group, and all observations equal to or above the cutoff point form the other group.

Value - This option is used only if you have a selected value as a cutoff point. The value entered in this field is used as a cutoff point. All observations falling below this value point form one group, and all observations equal to or above the cutoff point form the other group. If the data are already dichotomous, specify a numeric value that lies between those two values.

Sample output of runs test

```
RUN TEST: AGE  Age of the child
          Mean =      9.5424
Number of runs =      31
Expected number =     29.8136

Cases
    25  Lt Mean
    34  Ge Mean
-----
    59

          Z =      .3192          2-tailed P =      .7496
```

Sensitivity analysis

When using a questionnaire or an instrument to identify people who suffer from a specific disease (or any particular problem), the measure used is often associated with some diagnostic error. Two types of errors are possible: 1) the instrument can falsely classify a healthy subject as suffering from the disease (false positive), or 2) the instrument can fail to detect a person who has the disease (false negative). The error rates depend on the quality of the instrument and on the cut-off point used to classify the subjects. One problem is that increasing the cut-off point in order to reduce the number of false negatives will usually generate an increase in false positives. The sensitivity analysis allows one to assess the ability of a quantitative measure (X) to differentiate a dichotomous criterion condition (Y) and provide guidelines to choose a cut-off point that will offer an appropriate trade off between false-positive and false-negative error rates.

The program provides the level of sensitivity (i.e. the number of true cases detected by the test divided by the total number of true cases in the sample) and the specificity (i.e the number of true negatives divided by the total number of cases without the problem or disease) for each value of the quantitative measure. The program also displays, for each value of the test, the percentage of false-positives and false-negatives. If we plot, on a cartesian scale, and for each cut-off point the sensitivity at that point as a function of the proportion of false positives (or 1-specificity), and connect those points, we obtain what is known as a ROC (receiver operating characteristic) curve. This graph allows us to visualize the performance of the screening or diagnostic test used. If a test has no discriminatory value, the ROC curve will be a diagonal line with a 45 degree angle going from the lower left corner to the upper right corner of the graph. The ROC curve of a perfect test will be composed of a vertical line going from the lower left to the upper left point, and a horizontal line going from the upper left to the upper right corner. The higher the sensitivity and specificity of a test at each cutoff point, the closer the curve will be to the upper left corner of the graph, and the greater the area under the curve. The overall performance of a test can be quantified by computing the area under the ROC curve (AUC). A perfect test would yield an AUC of 1.0 while a useless test would yield an AUC of 0.5. Comparisons can also be made among alternative tests by comparing their ROC curves or their AUC values.

Various methods have been proposed for computing the area under the ROC curve. SIMSTAT provides two such methods. The first assumes that the categorical scale of the test results from an underlying continuous variable. The second method adopts a nonparametric strategy that does not make the assumption of a bivariate normal distribution. This measure of AUC is obtained by computing the area underneath the straight lines that connect the various observed points of the curves by using a trapezoidal method.

OPTIONS

Value - This option allows you to specify which value of the criterion variable will be used to identify positive diagnosis.

Scale orientation - This option allows you to specify whether the scale is positively or negatively related to the presence of the condition (disease). You must specify whether this condition is associated with higher or lower values of the scale.

Sensitivity statistics - This option allows you to obtain the level of sensitivity (proportion of positive cases correctly diagnosed as true) and specificity (proportion of negative cases effectively diagnosed as false) for each value of the quantitative measure. The values are reported in terms of absolute and relative frequencies.

Error rate statistics - This option allows you to obtain the number and percentage of false-positive and false-negative diagnoses for each value of the quantitative measure. The display also includes likelihood ratios for both positive and negative results.

ROC curve - This option allows you to obtain a receiver-operating-characteristic (ROC) curve that displays the relationship between the sensitivity and 1-specificity.

Error rate graph - This option provides a line chart that displays the evolution of the percentage of false-positives and false-negatives for the values of the scale.

Sample output of sensitivity analysis

SENSITIVITY ANALYSIS: DISORDER by TEST						
Area under ROC curve = .8286						
Cutting Point	Cases Included	Sensitivity	Control Included	Specificity	1-Specificity	
100.0	1	0.00	1	0.95	0.05	
9.0	2	0.10	1	0.95	0.05	
8.0	4	0.30	1	0.95	0.05	
7.0	9	0.60	3	0.86	0.14	
6.0	13	0.80	5	0.76	0.24	
5.0	18	0.90	9	0.57	0.43	
4.0	25	1.00	15	0.29	0.71	
3.0	29	1.00	19	0.10	0.90	
2.0	31	1.00	21	0.00	1.00	
Cutting Point	False Negatives	% False Negatives	False Positives	% False Positives	Pct False Results	Pos/Neg Ratio
100.0	10	100.0%	1	4.8%	35.5%	21.0
9.0	9	90.0%	1	4.8%	32.3%	18.9
8.0	7	70.0%	1	4.8%	25.8%	14.7
7.0	4	40.0%	3	14.3%	22.6%	2.8
6.0	2	20.0%	5	23.8%	22.6%	0.8
5.0	1	10.0%	9	42.9%	32.3%	0.2
4.0	0	0.0%	15	71.4%	48.4%	0.0
3.0	0	0.0%	19	90.5%	61.3%	0.0
2.0	0	0.0%	21	100.0%	67.7%	0.0

Sign test

The SIGN TEST tests the hypothesis that two variables have the same distribution. This is assessed by comparison of the numbers of positive and negative differences between values of the two variables.

OPTIONS

Direction - This option allows you to select either a one-tailed (directional) or two-tailed (non-directional) probability test.

Sample output of a sign test

```
SIGN TEST:  T2DEPRES With  T1DEPRES

      Cases
          27  -  Diffs (T2DEPRES Lt T1DEPRES)
          11  +  Diffs (T2DEPRES Gt T1DEPRES)
           2  Ties
          ---
          40  Total

      Z =  2.4333      1-tailed P =  .0075
```

Single-case design

Single case experimental design was originally developed for the study of animal and human behaviors in the context of controlled laboratory experiments. This methodology is now currently used in applied research such as the evaluation of behavior modification programs or other types of clinical interventions, the effects of pharmacological agents on behaviors, or the impact of educational and social intervention programs. It involves repeated objective measurement of a single subject (dependent variable) over a period of time interspersed with changes in the treatment condition (independent variable). If the application or withdrawal of the treatment condition is associated with systematic changes in the behavior of the subject, then it is inferred that the treatment has caused the observed changes. Whereas with traditional group design, variability is treated as error which should be controlled with the use of experimental design and statistical analysis in order to identify functional relations that supersede this variability, advocates of single-case experimental design stress the importance of identifying and controlling the source of this variability. By carefully monitoring the variability of measures taken on a single subject it becomes possible to identify the source of this variability and to develop effective methods of intervention. When applied to the evaluation of complex intervention programs, this method facilitates the identification of the active ingredients of the intervention. The reliance of single-case design on visual inspection (instead of statistical analysis) to identify the presence of a functional relationship also provides assurance that only potent interventions that produce clinically significant changes will be identified. The generality of the findings is strengthened through systematic replications of the original experiment with other subjects in various settings, conditions, or with other behaviors of the same subject.

The SINGLE CASE command provides some basic tools for studying the effect of an intervention on the behavior of a single subject. The procedure will display a graph representing the evolution of the dependent variable (Y) at various phases defined by the independent (X) variable. The dialog box allows one to obtain various statistics for each phase of the analysis as well as various graphic tools that can be used as judgemental aids to identify the experimental effect of the intervention (smoothed data, trend lines, control bars, etc.).

OPTIONS

Statistics - This option box allows you to display statistics about the different phases of the experiment. When set to Brief, the mean, standard deviation, minimum and maximum values, and the number of cases are displayed on a single line. To obtain additional statistics, such as the skewness, kurtosis, mode, median, etc., set this option to Detailed.

Cumulative frequency - This option allows you to represent the values of the series as a cumulative record of frequency.

Log transformation - This option converts all values in the series into their natural logarithms.

Trend line - This option allows you to draw, for each condition, a line representing either the mean, the regression slope, or a split middle trend line. This last method draws a line through the median values of the first and second halves of each series.

Smoothing - This option allows the application of two methods of identifying trends in noisy or irregular time series. The moving average procedure is obtained by averaging a selected number of points on either side of a target value, while the running median procedure is computed by finding the median value of a specified number of points on either side of the original value. Successive smoothing can be achieved by specifying more than one value. For example the following option:

```
Moving average 4 2 4
```

instructs SIMSTAT to proceed with three successive moving average smoothings using 4, 2, then 4 values to compute the mean.

Vertical line separators -This option lets you specify whether vertical lines should be used in the time-series chart to delineate the various phases of the experiment.

Control bars - This option allows you to superimpose 3 horizontal bars that represent the mean and the upper and lower limits of a confidence interval. Those bars can be used as a judgemental aid to identify a change in the series. The interval option allows you to specify the width of the confidence interval. Its value must be between 0% and 99%. The **minimum** and **maximum** options allow you to specify which observation in the series will be used to calculate the mean and the confidence limits. If those fields are left blank, the first and the last observations will be used. For example, entering 1 and 10 as the minimum and maximum values tells SIMSTAT to compute the mean and the confidence interval on the first 10 observations.

Confidence interval - This option allows you to set the confidence interval width for the control bars. This interval width is expressed as a percentage and must lie between 1% and 99%.

Time-series analysis

The TIME SERIES command allows the examination of time series data. The dialog box offers various transformations to remove trends or seasonal dependence in a series and provides a diagnostic of those transformations by displaying autocorrelation and partial autocorrelation function plots of the transformed series. This procedure also allows the application of two smoothing methods (i.e., moving average and running median) to identify trends in noisy time series data.

OPTIONS

ANALYSIS PAGE

Log transformation - When enabled, this option converts all values in the series into their natural logarithms.

Remove mean - When enabled, this option subtracts the mean from each value in the series.

Difference - This transformation subtracts the preceding value from each value in the series. The **Number** option allows you to set the number of difference operation to be performed on the series.

Seasonality - The Seasonality option allows you to remove the seasonality in a series by subtracting from every value the value that is a specified **Number** of lags behind it.

ACF plot - The ACF plot produces an autocorrelation function plot. This plot includes the autocorrelation value, its variance and probability, and a graphic (text mode) representation of those values from one to a specified number of lags.

PACF plot - This option allows you to obtain a table and a graphic representation of partial autocorrelations with their variances and probabilities.

Number of lags - This option allows you to specify the maximum number of lags to be displayed in the ACF and PACF plots.

CHART PAGE

Time series - This option allows the printing of a graphic representation of the transformed series in either text, graphics mode or both. The Number option allows you to restrict the number of observations to be plotted in text mode.

Smoothing - This option allows the application of two methods of identifying trends in noisy or irregular time series data. The Moving average procedure is obtained by averaging a selected number of points on either side of a target value, while the running median procedure is computed by finding the median value in a specified number of

points on either side of the original value. Successive smoothing can be achieved by specifying more than one value. For example the following option:

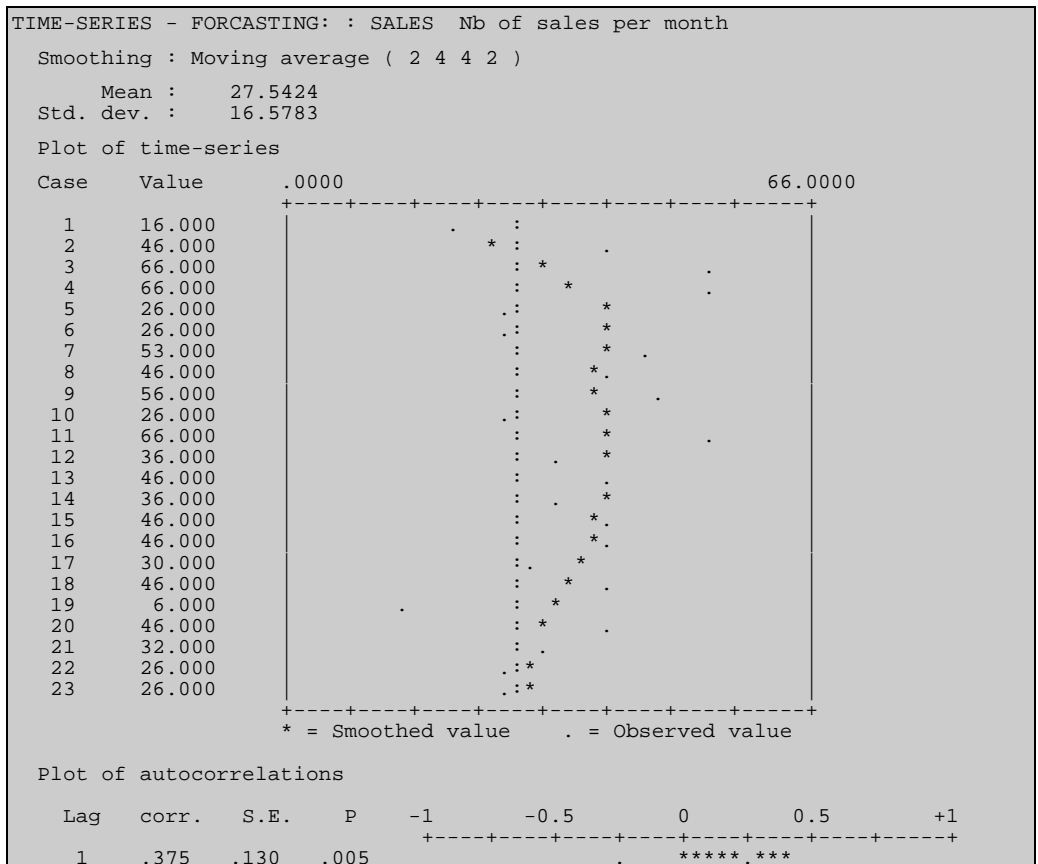
Moving average 4 2 4

instructs SIMSTAT to proceed with three successive moving average smoothings using 4, 2, then 4 values to compute the mean.

Control bars - This option allows you to display 3 horizontal bars that represent the mean and the upper and lower limits of a confidence interval. Those bars can be used as a judgemental aid to identify a change in the series. The **Minimum** and **Maximum** options allow you to specify on which observations in the series the mean and confidence limits will be calculated. If those fields are left blank, the first and the last observations will be used. For example, specifying the limits 1 and 10 tells SIMSTAT to compute the mean and the confidence interval for the first 10 observations.

Width - This option allows you to set the confidence interval width for the control bars. This interval width is expressed as a percentage and must lie between 1% and 99%.

Sample output of time series analysis



2	.399	.147	.009	.	*****	**
3	.255	.165	.127	.	*****	.
4	.359	.171	.040	.	*****	.
5	.379	.184	.043	.	*****	*
6	.296	.196	.137	.	*****	.
7	.318	.204	.124	.	*****	.
8	.266	.212	.214	.	*****	.
9	.348	.218	.115	.	*****	.
10	.104	.227	.647	.	***	.
11	.230	.228	.316	.	*****	.
12	.192	.232	.411	.	*****	.

Plot of partial autocorrelations

Lag	corr.	S.E.	P	-1	-0.5	0	0.5	+1
1	.375	.130	.005	+-----+	+-----+	+-----+	+-----+	+-----+
2	.301	.130	.024	.	.	*****	***	.
3	.048	.130	.714	.	.	**	.	.
4	.209	.130	.114	.	.	*****	.	.
5	.213	.130	.108	.	.	*****	.	.
6	.016	.130	.904	.	.	*	.	.
7	.092	.130	.484	.	.	***	.	.
8	.044	.130	.739	.	.	**	.	.
9	.115	.130	.379	.	.	***	.	.
10	-.229	.130	.084	.	.	*****	.	.
11	.045	.130	.730	.	.	**	.	.
12	.057	.130	.663	.	.	**	.	.

T-test analysis

T-TEST calculates either independent-sample t-tests or paired-sample t-tests to determine whether two sample means are significantly different.

The paired-sample (or correlated) t-test compares the means between each pair of variables assigned as dependent and independent. The independent t-test compares means on the dependent variable for two groups defined by values of the independent variable. SIMSTAT provides two distinct tests to take into account whether the two populations from which the samples are drawn have equal or unequal variances. You can also specify whether the null hypothesis should be evaluated using a one-tailed (directional) or a two-tailed (non-directional) test.

OPTIONS

ANALYSIS TAB

Type of design - This option determines whether the design includes paired (correlated) or independent samples.

Values of x - For independent samples, the program requires you to specify two values of the grouping (or independent) variable that will be used to define the two groups.

Direction - The Direction of the statistical test can be one-tailed (directional) or two-tailed (non-directional).

Confidence interval - The interval options allows you to set the width of the confidence intervals around the two measures of effect size. The interval width is expressed as a percentage between 1 and 99%. This value is also used when displaying barcharts with a confidence interval or an error bar graph representing a confidence interval around the mean.

CHART TAB

Mean/Error bar graph - This option displays a mean bar and/or an error bar representing the variability of the mean or of the values in each group.

Type of error - This option allows you to select whether the error bar will represent the standard deviation, the standard error or a user-defined confidence interval. The width of this interval is set by the interval option.

With barchart - This option allows you to draw bars where each bar represents the mean of a separate group. You can use the Type of Error option to add error bars.

Upper error bar only - When the bar chart option is selected and an error bar is requested, this option allows you to specify whether the error bars displayed with the mean bars should be displayed above and below the mean, or only above.

Link means - When chosen, this option connects the two means with a line.

Dual histogram - This option displays a dual histogram representing the distribution of two numeric variables. You must supply the **Number of bars** bars to plot. A **Normal curve** can also be superimposed on the histogram. The horizontal and vertical radio buttons allow you to select between horizontal histograms where the two charts are displayed side by side or vertical histograms displayed one above the other.

Sample output of an independent samples t-test

INDEPENDENT SAMPLES T-TEST							
AGGRESS by SEX		Nb of aggressive behavior Sex of the child					
		Number of Cases	Mean	Standard Deviation	Standard Error		
SEX	= 1	29	36.690	15.281	2.838		
SEX	= 2	30	18.700	12.636	2.307		
		Pooled Variance Estimate			Separate Variance Estimate		
F Value	1-tail Prob.	t Value	Degrees of Freedom	1-tail Prob.	t Value	Degrees of Freedom	1-tail Prob.
1.46	.314	4.94	57	.000	4.92	54.33	.000
Effect size		90.0% Confidence Interval					
R =	-.5471	[-.6827	To	-.3752]		
D =	-1.3073	[-1.868	To	-.8096]		

Sample output of a paired-sample t-test

PAIRED SAMPLES T-TEST: T2DEPRES by T1DEPRES								
Variable	Number of Cases	Mean	Standard Deviation	Standard Error				
T1DEPRES	38	37.895	6.849	1.111				
T2DEPRES	38	37.466	7.747	1.257				
Difference Mean	Standard Deviation	Standard Error	Corr.	1-tail Prob.	t Value	Degree of freedom	1-Tail Prob.	
.4291	6.7556	1.0959	.577	.000	.39	37	.349	
Effect size statistics								
Statistics		90.0% Confidence Interval						
r =	.0642	[-.2105	To	.3296]				
D =	.1288	[-.4307	To	.6982]				

Wilcoxon test

The WILCOXON matched-pairs signed-ranks test is a procedure used to test whether two related samples have been drawn from the same population. Like the sign test, it computes the difference between the values of the two variables but takes into account the magnitude as well as the direction of the differences. The Wilcoxon signed-ranks test is the nonparametric version of the t-test for paired samples.

OPTION

Direction - This option allows you to request the program to select either a one-tailed or two-tailed probability assessment.

Sample output of a Mann-Whitney U - Wilcoxon rank sum W test analysis

MANN-WHITNEY U - WILCOXON RANK SUM W TEST				
AGGRESS	Nb of aggressive behavior			
With SEX	Sex of the child			
Mean Rank	Cases			
40.48	29	SEX = 1.00		
19.87	30	SEX = 2.00		

	59	Total		
		EXACT	Corrected for ties	
U	W	1-tailed P	Z	1-tailed P
131.0	1174.0	.0000	-4.6325	.0000

7 - WORKING WITH CHARTS

Numerous statistical analyses within SIMSTAT include options to create high-resolution charts. This section describes the steps involved in the creation, modification, printing and saving of charts.

Creating specific charts

Rather than providing separate commands and modules for numerical and graphical analysis, SIMSTAT implements an integrated approach where charts are produced along with numerical analyses.

The table on the next page allows you to determine which statistical commands should be used to obtain a specific chart. To create a chart, select the variables and the proper statistical analysis command. Set the check box corresponding to the chart type you want. If some charting options are also displayed, adjust them to suit your needs.

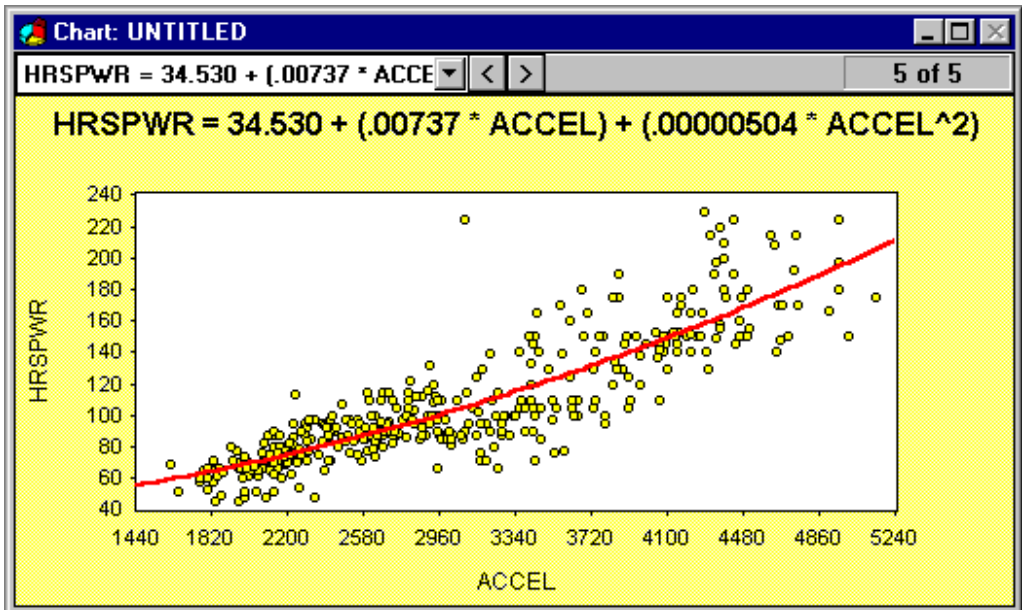
While the charting options available from statistical analysis dialog boxes are quite limited, it is possible to modify almost any chart attribute afterward. For more information on the various adjustments available, see section Customizing Charts on page 158.

Correspondence between chart types and statistical analysis commands


TYPE OF VARIABLES	TYPE OF CHART	COMMANDS
One variable (nominal)	Barchart	Descriptive Frequency
	Pie chart	Descriptive Frequency
	Pareto chart	Descriptive Frequency
One variable (metric)	Histogram	Descriptive Frequency
	Box-&-whisker plot	Descriptive Frequency
	Cumulative distribution chart (ogive)	Descriptive Frequency
	Normal probability plot	Descriptive Frequency
Two variables (nominal)	Barchart (2 dimensions)	Tables Crosstab
Two variables (metric x nominal)	Multiple box-&-whisker plot	Compare Breakdown
	Dual histogram	Compare T-test
	Error bars	Compare T-test
		Compare Oneway
	Mean bars	Compare T-test
		Compare Oneway
Two variables (metric x metric)	Deviation chart	Compare Oneway
	Scatterplot	Regression Regression
	Residual caseplot	Compare GLM Anova/Ancova
		Regression Regression
		Regression Multiple Regression
One metric variable over time	Residual scatterplot	Compare GLM Anova/Ancova
		Regression Regression
		Regression Multiple Regression
One metric variable over time	Time series plot	Time-series Time-series
	Autocorrelation plot (ACF & PACF)	Time-series Time-series
Metric variable over time x nominal	Interrupted time-series or single case experimental design	Time-series Single case
Other	ROC curve	Scale Sensitivity
	Error rate graph	Scale Sensitivity

Navigating in the chart window


All high-resolution charts created during a session are displayed in the Chart window. This window can be used to view the charts and perform various operations on individual charts or on the entire collection of charts. For example, you can modify the various chart attributes, save those charts to disk, export them to another application using the clipboard or disk files, or print them. It is also possible to delete a specific chart or to modify the order of those charts in this window.




TO DISPLAY THE PREVIOUS CHART

- Press the <Ctrl-P> key or click on the  icon.


TO DISPLAY THE NEXT CHART

- Press the <Ctrl-N> key or click on the  icon.

TO SELECT A SPECIFIC CHART FROM A LIST

- Click on the  button to display a list of all charts.
- Select the chart from the list.

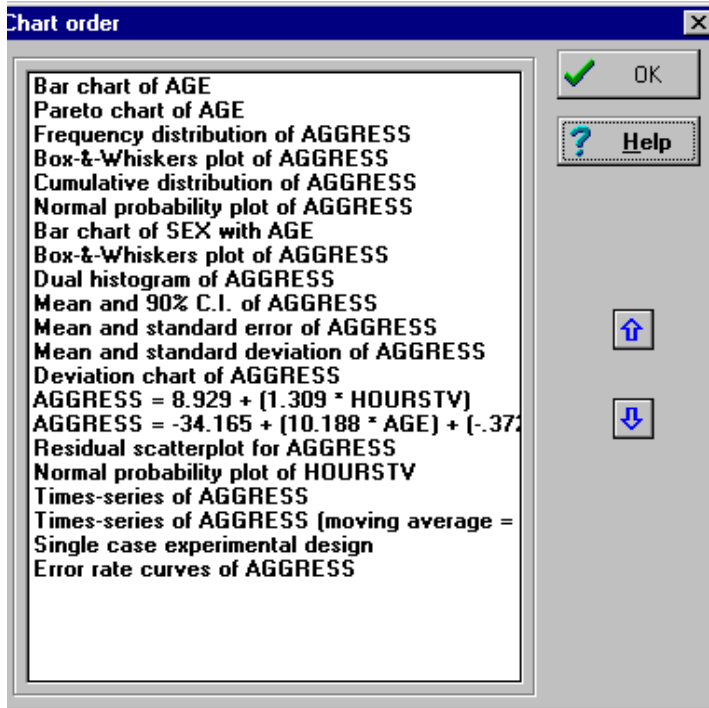
TO ERASE A SPECIFIC CHART

- Use the navigation keys or buttons to display the chart you want to delete.
- Select the ERASE command from the EDIT menu or click on the  button.

TO REORDER THE CHARTS IN THE CHART WINDOW

- Select the CHART ORDER command from the EDIT menu.

The following dialog box appears:





The list box displays the title of all charts in the Chart window.

- Select the chart you want to move by clicking on its title.
- Click on the up or down arrow buttons to move the chart up or down the list.


You may also drag the chart to its new location by keeping the mouse button down and dragging the chart title to its new location.


TO WRITE THE CHARTS TO DISK

- Select the CHART | SAVE command from the FILE menu or click on the  button while the Chart window is active.
- If the charts have never been saved before, a File Save dialog box will appear.
- Enter the name of the file under which you want to save the charts and press <Enter> or click on the **OK** button. By default, SIMSTAT uses the .CHX extension for chart files. If no extension is given, the program automatically adds this extension to the end of the file name.


 To save the charts under a different file name, choose the CHART | SAVE AS command and provide a new file name.

TO RETRIEVE CHARTS STORED ON DISK


- Select the CHART | OPEN command from the FILE menu or click on the  button while the Chart window is active.

 Rather than creating a new chart file, you may want to add new charts to an existing chart file. To do this, open the existing chart file where you want the new charts to be placed before creating those charts.

TO MERGE TWO SEPARATE CHART FILES

- Open the first chart file containing the charts that should be positioned at the beginning of the new file by using the CHART | OPEN command from the FILE menu, or by clicking on the  button while the Chart window is active.
- Select the CHART | APPEND FROM command from the FILE menu, and select the chart file that should be added to the end of the currently opened chart file.

TO CLEAR THE CHART WINDOW

- Select the CHART | NEW command from the FILE menu or click on the  button while the Chart window is active.

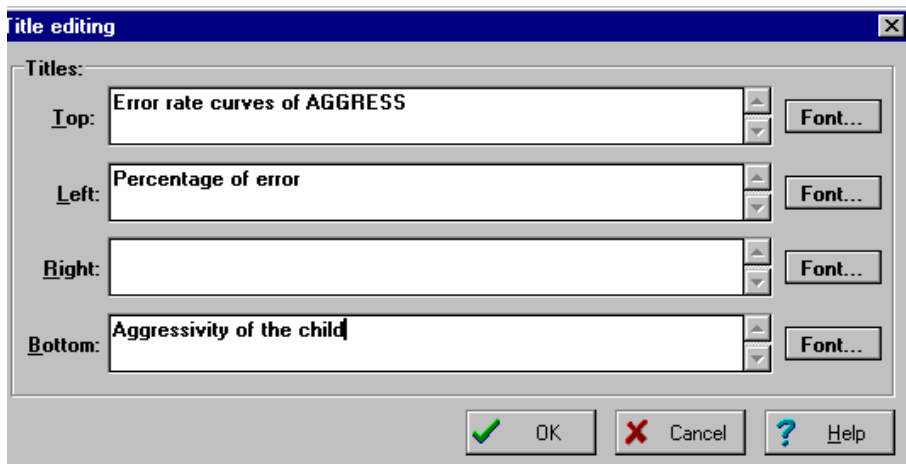
If any modification has been made to a chart in the current Chart window or if new charts have been created, you will be asked if you would like to save the modifications to disk. Select **Yes** if you want to save those modifications or **No** to clear the chart window without saving the changes.

Customizing charts

After you create one or several charts, you can edit their title and axis labels, choose a different font or font size, adjust the scaling on either axis, experiment with different colors or patterns, add a legend, or make other adjustments to suit your needs. The current section presents the various options available for customizing charts.

Editing titles and axis labels

Proper titles and axis labels are of utmost importance when describing the information displayed in a chart. By default, SIMSTAT will use variable names and labels as well as some predefined strings to provide such descriptions. To add or edit those titles, select the TITLES command from the CHART menu. The following dialog box will appear:



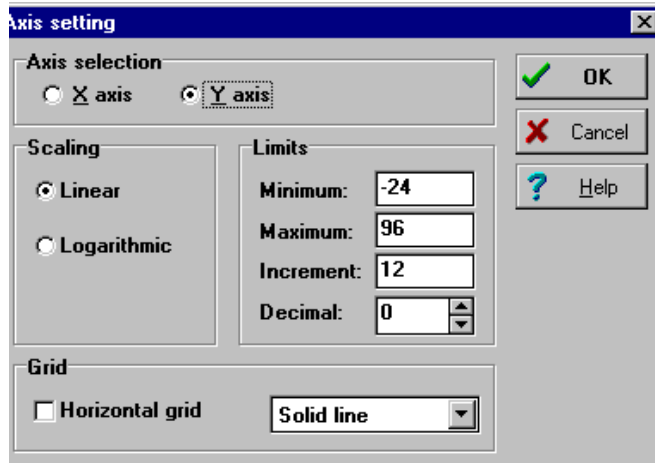
This dialog box provides 4 edit fields where you can create or edit the existing chart title, and the labels on the left, bottom and right axis.

You can enter several lines of text for each title by pressing the <Enter> key at the end of a line before entering the next line.

Font buttons on the right side of each edit box allow you to quickly change the font size or style of the related title (Please note that the font setting is a global option and will be applied to all charts in the Chart windows).

Editing axis scaling and grid

The **AXIS** command on the **CHART** menu allows you to specify various axis options such as the axis limits, increments and the number of decimal places used to display axis values. You can also control the display of horizontal and vertical grids. When this command is invoked, the following dialog box appears:



Axis selection - This group of radio buttons allows you to specify the appropriate axis that you want to customize. In some charts, only the Y axis will be available.

Linear or Logarithmic scale - In some types of chart you can choose between a linear or logarithmic axis. In a linear scale, the value of each major division is exactly the same. In a logarithmically scaled axis, each major division of the axis represents 10 times the value of the previous major division.

Minimum and Maximum - SIMSTAT automatically adjusts the axis scales to fit the range of values plotted against it. To manually set these values, type the desired minimum and maximum for the axis selected.

Increment - SIMSTAT automatically selects the initial increment value used for the axis. By default, this increment value is set to display 10 tick marks per axis. Increasing or decreasing this value affects the distance between these tick marks as well as their number. Grid lines are also affected by modification of this value.

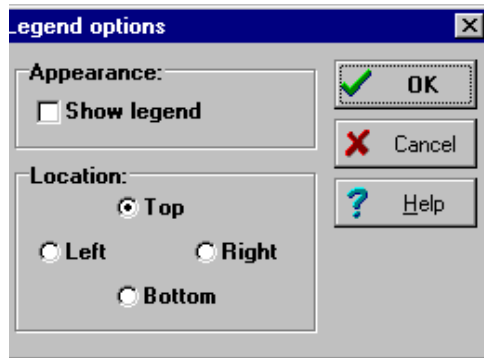
Decimal - This option allows you to increase or decrease the number of decimal places used to display the values on the axis.

Grid lines - This option lets you turn horizontal (Y axis) and vertical (X axis) grid lines on and off. Grid lines extend from each tick mark on an axis to the opposite side of the graph. To increase or decrease the number of grid lines per axis or the distance between those lines, change the Increment value of the current scale.

A list box also allows you to choose among 3 different line styles to draw those grid lines.

Controlling the legend display

This command brings a dialog box that allows one to specify whether a legend should be displayed on the chart and to set the legend position.



To resize the width or height of a legend, simply drag its border to resize it.

To set a default legend location upon creation of new charts, see the GLOBAL OPTIONS command.

Displaying data point values

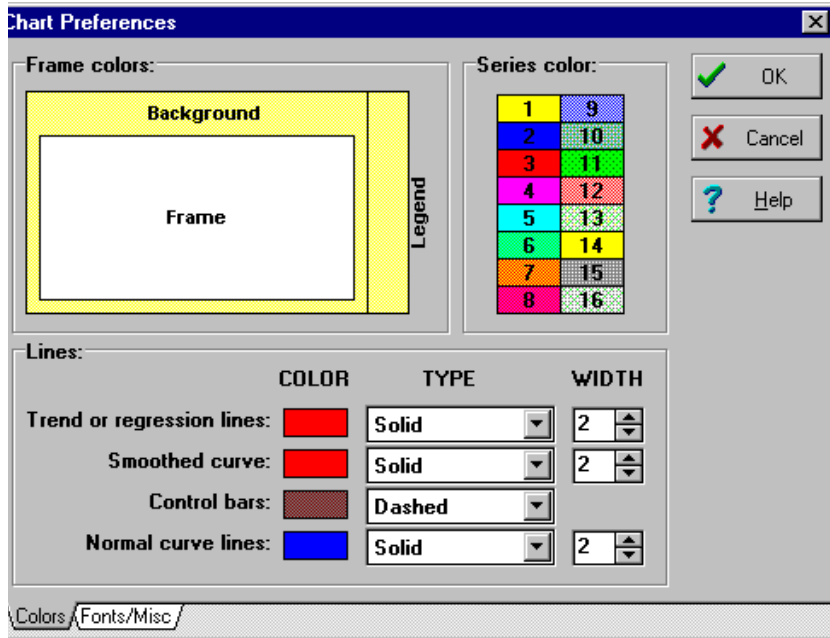
Some chart types give you the option of displaying the numeric value associated with each data point. When such an option is available, the DATA LABELS command in the CHART menu is enabled. Selecting this command immediately displays those data labels. To remove the data labels, select the DATA LABELS command again.

Modifying specific chart options

The SPECIFIC OPTIONS dialog boxes control the display of properties that are specific to the currently displayed chart type. For most chart types, the options available in this dialog box are the same as those displayed in the dialog box used to create the charts. For example, if the current chart is a histogram, this dialog box gives you the option of removing or adding the normal distribution curve, or changing the number of intervals.

Setting global options

Selecting the GLOBAL OPTIONS command from the CHART pull-down menu displays a multi-page dialog box containing options that are shared by all charts. Any change made to these options will affect all charts in the current Chart window. These options include frame and series colors, series markers, titles and label font sizes and styles.



Adjusting the color of various chart components

SIMSTAT allows you to change the color of various chart components such as the chart itself, the frame and legend backgrounds, the series, or the reference lines drawn over some chart types. To change the color of an item:

- Click on this item in the Chart Preferences dialog box. This will display a standard color dialog box.
- Select the new color and click on the **OK** button to leave the Color dialog box.

The Color Scheme option on the next page of the Chart Preferences dialog box allows you to use either colors, patterns or both to differentiate the various series or bars in the chart.

Adjusting line style and width

You can change the line style of reference lines by clicking on the drop down arrow key of the proper Type list box and selecting the line style you want. To hide a reference line, select **None** as the new line style.

When a solid line style is chosen, you can use the Width spin button to set the thickness of the line to a value between 0 and 9.

Setting the font size and style of charts components

To change the title font, the axis labels or the value labels:

- Select the second page of the Chart Preferences dialog box by clicking on the Fonts/Misc tab,
- In the String list box, select the text component you want to modify by clicking on its name,
- Change the font name, size, color or style.

Miscellaneous chart options


Color scheme - This option allows you to use either colors, patterns or both colors and patterns to differentiate the various series or bars in a chart.


3-D frame - When this option is enabled, SIMSTAT draws a 3D frame.

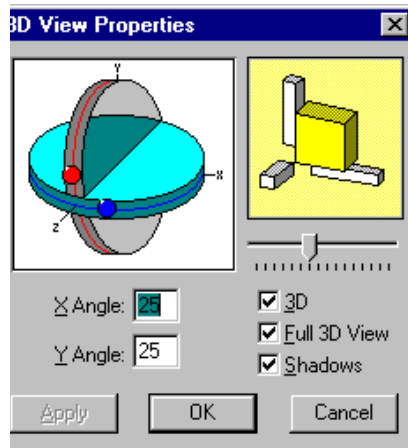
Point marker size - This option lets you specify the size of point markers for all series of a chart. Marker size must be set between 1 and 20.

Default legend location - This group of options lets you specify whether to display a legend when a new chart is created, and its default location. These options apply to piecharts, bar charts, and Pareto charts.

3D View

Some charts can be displayed using a 3D perspective. When such an option is available, the  button on the toolbar is enabled. You can click on this button to turn on/off the 3D perspective for the current chart.

Selecting the 3D VIEW command from the CHART menu or clicking on the  button gives access to the 3D View Property dialog box. This allows you to adjust the viewing angles, object depth, and shadows of the 3D charts.



To adjust the viewing angles

- Check the Full 3D View box.
- Drag the marbles to the desired angles, or type the desired angles in the fields.
- Verify the desired view shown in the sample rotation frame.
- Click on the **Apply** button to modify the angles in the background chart while remaining in the rotation dialog, or click on the **OK** button to apply these angles and exit the rotation dialog box..

To control the 3D depth

You can also control the depth of the 3D chart with the sliding control located underneath the rotation preview. When you drag the sliding control to the right you will increase the depth (in comparison to the chart width). Sliding it to the left will decrease the chart depth.


To alter the 3D shadows

In this dialog box, you'll find the shadows check box which will allow you to turn the shadow for the markers in 3D mode on and off. Turning it off will cause the back side of the markers to paint in the same color assigned to the front of the marker.

Zooming in and out

Sometimes, you may want to zoom on a chart to view it in more detail. While it is possible to do this by adjusting the axis scaling, SIMSTAT provides a more convenient method using the mouse.

To zoom in on a specific portion of a chart:

- Activate the zooming feature by clicking on the  button or selecting the ZOOM IN command from the CHART menu.
- Click with the mouse on the upper left corner of the area you want to display.
- Drag the mouse cursor to the lower left corner of the area.
- Release the mouse button.

SIMSTAT automatically resets the axis minimum and maximum values as well as the increment value to coordinates near the defined area. Those new axis limits are usually saved as soon as you move to another chart or switch to another window.

To restore the original axis scaling

- Select the ZOOM RESET command from the CHART menu to revert the effect of one or several ZOOM IN commands and restore the original axis scaling of a chart.

This command should be used immediately after a ZOOM IN command has been performed because any other modification to the chart or any window operation may cause the new axis limits to be saved, resulting in the loss of the original scaling information.

Exporting charts

SIMSTAT lets you transfer charts to the clipboard, or export them to other formats, so that they can be viewed or edited by other applications. SIMSTAT currently supports 3 different file formats: Windows Metafiles, Windows Bitmap, and tab separated values files.

Window metafile format (.WMF)

The Windows Metafile format stores graphic images in a vector based format. The resulting file is much smaller than a bitmap format, and the resolution of the graphic image becomes device independent. In other words, its resolution will be determined by the output device. This file format may be inserted in most Windows word processing program, modified using many drawing packages (such as Corel Draw or Microsoft PowerPoint) or paint programs (e.g., Microsoft Paint).

Windows bitmap format (.BMP)

Charts saved in a bitmap format are stored as series of pixels of different colors. The charts are stored as they are displayed on the screen. Therefore, their resolution is determined by the resolution of the screen as well as the size of the graph. This means that greater resolution can be achieved by maximizing the Chart Window before exporting the chart to disk or to the clipboard. However, you should be aware that bitmap files take a lot of disk space, and increasing the size of a bitmap image can increase its required disk space by several hundred kilobytes.

Like the Metafile format, bitmap files may be imported into most Windows word processing, drawing or painting programs.

Tab separated values format (.TAB)

Rather than exporting charts in a graphic format, you may prefer to extract the values of a chart and use those values in another charting program (such as Harvard Graphics, Chart XL., Microsoft PowerPoint), or a spreadsheet program with advanced charting features. Exporting a chart to a tab separated values format produces a generic text document that can be pasted into or imported by a wide range of charting and spreadsheet applications.

TO TRANSFER A SINGLE CHART TO THE CLIPBOARD

- Display in the Chart window the chart you want to transfer.
- Select the COPY AS command from the EDIT menu.
- Select the format in which you want to transfer the chart.

TO EXPORT A SINGLE CHART TO FILE

- Display in the Chart window the chart you want to export.
- Select the CHART | EXPORT command from the FILE menu.
- Activate the Current Chart radio button to specify that only the current chart should be exported to disk.
- Set the List of File Type list box to the format in which you want to export the current chart.
- Enter a valid file name and click **OK** to exit the dialog box and export the chart.

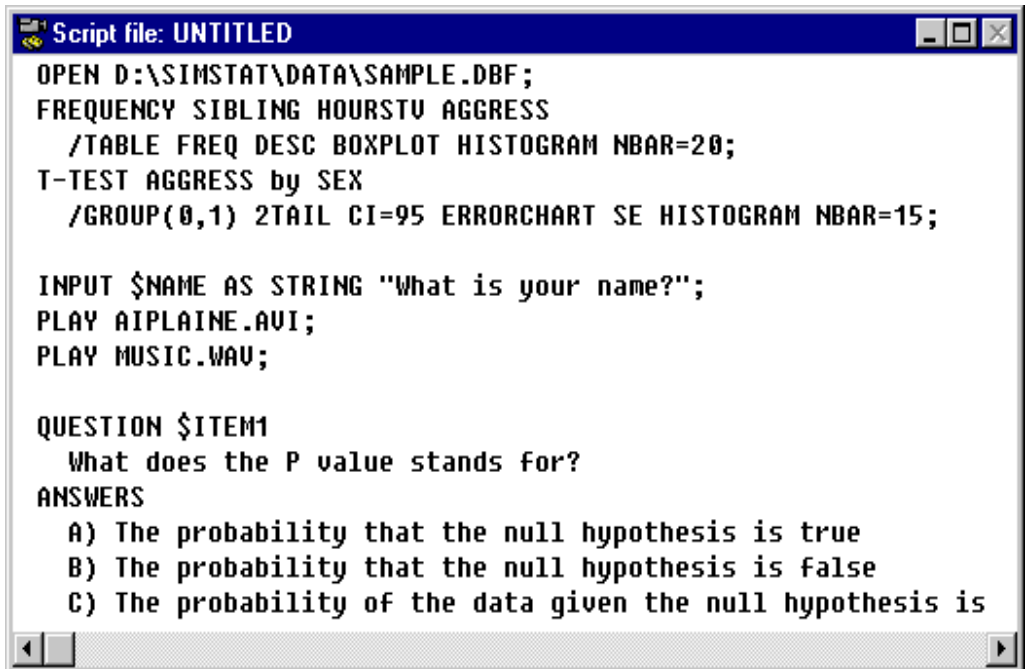
TO EXPORT ALL CHARTS IN THE CHART WINDOW TO FILES

- Make the Chart window active.
- Select the CHART | EXPORT command from the FILE menu.
- Activate the All Charts radio button.
- Set the List of File Type list box to the format in which you want to export the charts.
- Enter a valid prefix that will be used to create file names and click **OK** to exit the dialog box and export the charts. This prefix can have a maximum length of 5 characters. A serial number between 001 and 999 will be automatically added to this prefix to create unique file names.

NOTE: Great care should be taken when exporting several charts in .BMP format since this single command may produce several megabytes of bitmap files.

8 - USING SCRIPTS

The script window is used to enter, edit, and execute commands. Those commands can be read from a script file on disk, typed in by the user or automatically generated by the program. When used with the RECORD feature, the script window can also be used as a log window to keep track of the analyses performed during a session. Those commands may then be executed again, providing an efficient way to automate statistical analysis. Additional commands also allow one to create demonstration programs, computer assisted teaching lessons, and even computer assisted data entry programs.



```

Script file: UNTITLED
OPEN D:\SIMSTAT\DATA\SAMPLE.DBF;
FREQUENCY SIBLING HOURSTU AGGRESS
  /TABLE FREQ DESC BOXPLOT HISTOGRAM NBAR=20;
T-TEST AGGRESS by SEX
  /GROUP(0,1) 2TAIL CI=95 ERRORCHART SE HISTOGRAM NBAR=15;

INPUT $NAME AS STRING "What is your name?";
PLAY AIPLAINE.AVI;
PLAY MUSIC.WAV;

QUESTION $ITEM1
  What does the P value stands for?
ANSWERS
  A) The probability that the null hypothesis is true
  B) The probability that the null hypothesis is false
  C) The probability of the data given the null hypothesis is

```

This section introduces you to various tasks that may be performed with script files such as how to:

- Open an existing script file.
- Navigate in the script window and edit its contents.
- Execute a script or only part of it.
- Use the RECORD SCRIPT command to automatically generate commands.
- Write script files to disk.

The final part of this chapter provides programming information, including a reference section with a description of all available commands, their syntax and related options.

Introduction to the scripting feature

While SIMSTAT pull-down menus and open panels allow you to do almost everything you need, it is also possible to perform analyses using the SIMSTAT batch command language. But, why should you bother using this script language if you can do everything you need with the user interface?

First, using the script language allows you to keep track of what you did during a session. This may prove very useful if you want to come back later and inspect what you did, or if you have to provide to someone else a detailed description of your analyses.

The second advantage is that it allows you to automate statistical processing of your data files. Often, you will have to perform an identical series of analyses on the same data file on several occasions, such as every month or every year. It may become easier and faster to resubmit the same script file than to remember all you did and do it again with the menus and dialog boxes. You may also want to perform similar analyses on different data files. In that situation, modifying an existing script file may be more efficient than starting from scratch with the menus.

Another reason to use the script language is that you may want to write sets of procedures that would be executed by someone else less familiar with statistics or with the operation of SIMSTAT.

But SIMSTAT's script language goes beyond the simple automation of statistical analysis and provides some commands that allow you to write interactive tutorials, demonstration programs or even small applications to be used by someone else. These special commands allow you to display textual information, wait for a key to be pressed, ask questions, construct bouncing bar menus, play music or sound files, display graphics or animations, etc.

Using normal and encrypted script files

SIMSTAT normal script files are plain text files with an .SCR extension. They are usually created and edited from within the program but may also be created or edited using almost any word processor or text editing program. However, if you use a word processor, make sure that you save the script file as a plain text file.

SIMSTAT can also execute encrypted and compressed script files with an .SCZ extension. Once you have developed a program, you may want to prevent others from altering your source file or simply hide its contents. One reason would be to make sure that no one else will commercialize your entire script or parts of it under their own name. This encrypting feature may also be useful to prevent unauthorized changes to the original program or, in the context of computer assisted instruction, to prevent students from cheating by looking at your code. Encrypted files are created quite simply by saving an opened script under a filename with a .SCZ extension. The resulting file will be about 50% to 80% smaller than the original file and may be run from within SIMSTAT just as any other script file. However, the file can no longer be viewed or edited either from within SIMSTAT or from an external editor.

Opening an existing script

To open an existing script file, select the **SCRIPT | OPEN** command from the **FILE** menu. This evokes an Open File dialog box. When this dialog box is displayed, the program points to the default data directory and displays all available data files in this directory in the File Name list box. To open a file, double click on its name or select it and click on the **OK** button.

If the name of the script file you want to open is not displayed, type the filename in the File Name box (including drive and path if necessary) and select the **OK** button.


You may also use the following methods to locate the data file:

- If the script file is on a different disk, click on the down arrow of the Drives list box to display available drives and select the disk where the file is located.
- If the file is in a different directory, double-click on the directory names in the Directory (or Folder) to move through the directory tree.

If the file name is displayed in the File Name list box, double-click it to open the file or select it and click on the **OK** button.

If you want to open a script file that has been used previously, click on the down arrow button at the right side of the File Name edit box and select the filename.

If the selected file is an encrypted file (with a **.SCZ** extension), the text editor will be hidden and only the name of file will be displayed in the middle of the script window.

-  To prevent accidental modifications to the contents of a script file, activate the Read Only check box in the Open File dialog box before clicking on the **OK** button. While this procedure disables all editing features including the **RECORD SCRIPT** command, it will still be possible to view or print the content of the script window, and to cut text from it and paste this text to another **SIMSTAT** window or another application. To prevent these operations, create an encrypted version of the script file by saving the file under a new file name with a **.SCZ** extension.

Navigating in the script window

To move the caret to a specific location, click on that location with the mouse. You can also use the following keys to navigate in the script window:

Key	Action
<Left>	Move the caret one character to the left.
<Right>	Move the caret one character to the right.
<Up>	Move one line up.
<Down>	Move one line down.
<PgUp>	Move one screen up.
<PgDn>	Move one screen down.
<Home>	Move to the beginning of the current line.
<End>	Move to the end of the current line.
<Ctrl-Home>	Move to the first line of the script.
<Ctrl-End>	Move to the last line of the script.
<Ctrl-Right>	Move to the beginning of the next word.
<Ctrl-Left>	Move to the beginning of the previous word.

You can also move to a specific string within the current script by choosing the FIND command in the EDIT menu or by pressing <Ctrl-F>.

The followings editing commands are also available:

Key	Action
<BackSpace>	Delete the character to the left of the caret.
	Delete the current character or selected text.
<Ctrl-Ins> or <Ctrl-C>	Copy the selected text to the clipboard.
<Shift-Del> or <Ctrl-X>	Delete the selected text after copying it to the clipboard.
<Shift-Ins> or <Ctrl-V>	Paste the text from the clipboard.
<Ctrl-Z>	Undo the last operation.
<Ctrl-Shift-0> to <Ctrl-Shift-9>	Set the position of marker (0-9) to the current caret position.
<Ctrl-0> to <Ctrl-9>	Move to the previously set marker (0-9).

Using the RECORD SCRIPT Feature

This feature automatically generates proper commands corresponding to the action you undertake using the menus and dialog boxes, and appends them to the end of the current script.

To activate this feature, select the RECORD option from the SCRIPT menu. The **Record** keyword will appear on the status line. From now on, almost every action you perform will be recorded by SIMSTAT. To deactivate this feature, follow the same steps a second time. You can also press the <Ctrl-F10> key combination or click on the **Record** keyword on the status line to toggle the RECORD script feature on and off.


The extensive correspondence between the commands/keywords and the options available through the dialog boxes greatly facilitates the learning of the script language syntax. However, the easiest way for a beginner to write but also to become familiar with this syntax and the various keywords is to use the RECORD SCRIPT feature. You can experiment by performing some analysis and looking closely at the commands generated. This feature is also an efficient method to write script files rapidly and easily.


Running a script

To run an entire script


- Click on the  button or select the RUN command from the SCRIPT menu.

To run a specific block of commands

- Select the commands that you want to execute using either the mouse or the keyboard. Make sure that the beginning of the selected text is located on the first line of a valid command.
- Click on the  button or select the RUN SELECTION command from the SCRIPT menu.

 To run a single command, simply position the caret anywhere on the first line of the command you want to execute and activate the RUN SELECTION command.

To start the execution of a script at a specific line

- Position the cursor anywhere on the line where you want SIMSTAT to begin the execution of commands.
- Click on the  button or select the RUN FROM CURSOR command from the SCRIPT menu.

Saving script files

To save a script file, choose the **SCRIPT | SAVE** command from the **FILE** menu. A file save dialog box appears. By default, SIMSTAT uses the **.SCR** extension for script files. If no extension is given, the program automatically adds this extension to the end of the file name.

To save the contents of the script window in an encrypted and compressed file, enter a file name with a **.SCZ** extension or select the **Encrypted Script (.SCZ)** option from the **Save File as Type** list box and enter a valid file name without an extension.

9- SCRIPT LANGUAGE REFERENCE

Syntax Convention

This section outlines the syntax conventions of the various commands and options. Unless specified otherwise, you can type commands and options in either uppercase or lowercase letters. You will find below a short description of those elements.

UPPERCASE

Items in capital letters are keywords. Keywords are a required part of the statement syntax, unless they are enclosed in brackets or specified as optional.

lowercase italic

Items in lowercase italic characters are placeholders for information you must supply in the statement. Several types of information can be required such as:

variable a single variable name.

varlist One or several variable names. A set of consecutive variables can be designated by typing the first and last variable names separated by two dots (..). For example the DEPRES1..DEPRES29 expression refers to all the variables in the active file starting from DEPRES1 up to, and including DEPRES29. A variable list can span over several lines.

filename A filename with a valid extension. By default the file is assumed to reside in the starting directory. To refer to a file in another location, specify the full path name.

integer Integer value. You can either use an equal sign '=' between the option and the integer or put the integer between parentheses.

real Real value. May be entered in either normal or scientific notation, and can be put after an equal sign or between parentheses.

string A string of alphabetical as well as numeric characters. Some commands require the string to be enclosed between quotation marks (").

color A keyword representing a color. Valid keywords are:

BLACK	PURPLE	BLUE
MAROON	TEAL	FUSCHIA
GREEN	GRAY	AQUA
OLIVE	SILVER	WHITE
NAVY	RED	
BLUE	LIME	

[]

Items inside square brackets are optional.

	A vertical bar indicates a choice between two or more items.
item, item, ...	A horizontal three-dot ellipsis means more of the preceding items can be used in a single-line statement.
keyword	A vertical three-dot ellipsis is used to indicate block-structured statements. Textual information can be put between the beginning and end of the block.
:	
ending keyword	
*	Comments can be inserted anywhere in the file by placing an asterisk as the first character.
{ }	Comments can also be included anywhere within a command by enclosing them between braces.
PANEL	The PANEL keyword can be used with almost any statistical analysis command to display the dialog box associated with the command, allowing the user to change options before performing the analysis.
;	The ';' character is used as a command terminator. Each command must end with the ';' character.
:	A colon at the beginning of a line indicates the presence of a label (see the GOTO and GOSUB command).
/	The slash '/' character is used to separate the various options from the commands and the variables. While only one slash is required, it is also possible to include more slashes to visually separate various groups of options.
multiple lines	It is always possible to spread a long command over several lines, or even insert blank lines within a command. The semi-colon ';' character is always used to indicate the end of the command.

Using memory and data file variables

A script can contain 2 different kinds of variables: Memory and data file variables.

A memory variable is the name of a location in memory that stores a value. The value of a memory variable can change during script execution. Every variable has a name that begins with a dollar sign as its first character, a data type (i.e., numeric or alphanumeric), and a value.

To declare a memory variable, you must provide its name and data type. The explicit declaration of a variable is done with a DIM statement. For example, to declare an alphanumeric variable called \$NAME and a numeric variable called \$AGE, you must enter the following commands:

```
DIM $NAME AS STRING;  
DIM $AGE AS NUMERIC;
```

You may also declare a new memory variable in any other command by explicitly stating its data type on its first appearance. For example:

```
DIM $AGE AS NUMERIC;
LET $AGE = 12;
```

can also be expressed as:

```
LET $AGE AS NUMERIC = 12;
```

SIMSTAT gives each variable an initial value at the time it is declared. A string variable is initialized to the empty string, a string with no characters (""). A numeric variable is initialized to zero.

You can also access any variable (or field) of the currently opened data file by putting a DB. prefix to the variable's name. For example, to modify the value of a variable named AGE in the current data file, you can use a LET command in the following way:

```
LET DB.AGE = 12;
```

You can also read the value stored in this variable just like you would do with any other memory variable:

```
IF DB.AGE < 18 THEN LET CATEGORY = 1;
```

All reading and writing operations on data file variables are performed on the currently selected record. In order to access the various records in the data file you need to use the RECORD command.

SIMSTAT provides several predefined memory variables holding time and date related information:

Variable	Returns
\$CURRENT_YEAR	Current year
\$CURRENT_MONTH	Current month number (from 1 to 12)
\$CURRENT_DAY	Current day of the month (from 1 to 31)
\$CURRENT_WEEKDAY	Current day of the week (from 1 to 7)
\$CURRENT_TIME	Current time of the day in seconds with one decimal place (tenth of a second).
\$CURRENT_RECORD	Current record number
\$NB_RECORDS	Number of records currently displayed

Expression Operators and Functions

Some script commands such as LET or IF may include expressions where basic arithmetic, trigonometric, or random number operations are performed. Those expressions may contain the following operators or functions:

ARITHMETIC OPERATOR

+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponentiation

NUMERIC AND TRIGONOMETRIC FUNCTIONS

Syntax FUNCTION (value, variable or expression)

ABS ()	Absolute
ACOS ()	ArcCosine
ASIN ()	ArcSine
ATAN ()	Arctangent
CSC ()	Cosecant
COS ()	Cosine
EXP ()	Exponential
FACT ()	Factorial
LN ()	Natural logarithm
LOG ()	Base-10 logarithm
MOD10 ()	Modulus
RND ()	Round
SEC ()	Secant
SQRT ()	Square root
SQR ()	Square
SIN ()	Sine
TAN ()	Tangent
TRUNC ()	Truncate

RANDOM NUMBER FUNCTIONS

Syntax FUNCTION (value, variable or expression)

NORMAL(**x**) Pseudo-random number generated from a normal distribution with mean of 0 and standard deviation of X.

UNIFORM(x) Pseudo-random number generated from a uniform distribution between 0 and X.

One line descriptions of commands

You will find below a short description of all available script commands grouped in four broad categories. For more detailed information on each of these commands, refer to the description available in the next section where commands are presented alphabetically.

Database commands:

APPENDFROM	Appends data in a .DBF file to the current data file.
COMPUTE	Stores the result of a computation in a new variable or an existing one.
DATA...ENDDATA	Creates a temporary data file.
FILTER	Filters the records.
IMPORT	Imports data from another file format.
OPEN	Opens a data file, a notebook file or a chart file.
RANK	Transforms the values of a variable into ranks.
RECODE	Recodes the numeric values of a variable.
RECORD	Moves within the records of the data file, or adds a new record.
SORT	Sorts records on one or several variables.
VARDEF	Defines variables (label, missing values, display width & decimals, etc.).
VLABELS	Defines value labels.
WEIGHT	Weights records using a variable.

Flow control commands:

CALL	Runs another script program and returns.
GOTO	Branches to another part of the program.
GOSUB	Branches to and returns from a subroutine.
IF...THEN...ELSE	Carries out a command based on a specified condition, otherwise performs another command.
DEFINE or DIM	Defines a new memory variable.
DELAY	Temporarily stops the program execution for a specific length of time.
LET	Assigns a value or the result of an expression to a variable.
QUIT	Quits SIMSTAT.
RUN	Runs an external DOS or Windows program.
STOP	Stops the script and returns to the SIMSTAT menu.

Interactive commands:

BOX...ENDBOX	Displays a dialog box with textual information.
BEEP	Generates a beep sound.
INPUT	Gets a numeric or a string value from the user.
MENU...END	Creates and displays a bouncing bar menu.
QBOX	Displays a dialog box with textual information
QUESTION...	Displays a multiple items question dialog box.

Statistical analysis commands:

BINOMIAL	Binomial test
BOOTSTRAP1	Univariate bootstrap analysis
BOOTSTRAP2	Bivariate bootstrap analysis
BREAKDOWN	Breakdown analysis
CHISQUARE	Chi-square one sample test
CLUSTER	Cluster analysis (requires MVSP v2.2)
CORANAL	Correspondence analysis (requires MVSP v2.2)
CORRELATION	Correlation matrix
CROSSTAB	Contingency crosstabulation
DESC	Descriptive analysis
DIVERSITY	Diversity Index computation (requires MVSP v2.2)
FACTOR	Factor analysis (requires EFA v3.0)
FULL	Full analysis on bootstrap samples
FREQUENCY	Frequency analysis
FRIEDMAN	Friedman test
GLMANOVA	GLM analysis of variance and covariance
INTERRATERS	Interraters crosstabulation
ITEM	Classical item analysis (requires StatItem v1.0)
KRUSKAL	Kruskal-Wallis Anova
KS1	Kolmogorov-Smirnov one sample test
KS2	Kolmogorov-Smirnov two sample test
LIST	Listing of data
LOGISTIC	Logistic regression (requires Logistic v3.11)
MANN	Mann-Whitney U, Wilcoxon W test
MCNEMAR	McNemar test
MEDIAN	Median test
MOSES	Moses test of extreme reactions
MRESPONSE	Multiple responses analysis
MULTREG	Multiple regression analysis
NPAR	Nonparametric association matrix
ONEWAY	Oneway analysis of variance
PCA	Principal component analysis (requires MVSP v2.2)
PCO	Principal correspondence analysis (requires MVSP v2.2)
REGRESSION	Linear & nonlinear regression analysis
RELIABILITY	Reliability analysis
SCED	Single case experimental design
RANDOM	Random bivariate bootstrap analysis
RUNS	Runs test
SENSITIVITY	Sensitivity analysis (ROC curves)
SIGN	Sign test
T-TEST	T-Test (independent and paired)
TIME-SERIES	Time-series analysis
WILCOXON	Wilcoxon signed rank test

Environment and Multimedia control commands:

CHART	Controls the appearance or display of the current chart.
NOTE	Displays a text string at the bottom of the screen.
PLAY	Plays a WAV sound file, a MDI music file, or an AVI movie file.
PICTURE	Displays a bitmap file on screen.
PRINT	Prints the contents of a window.
SAVE	Saves the contents of a window on disk.
SCREEN	Creates a blank background screen.
TITLE	Displays a text string at the top of the screen.
WINDOW	Controls the size and location of any SIMSTAT window.

Command descriptions

APPENDFROM

Syntax:

```
APPENDFROM filename;
```

Description:

The APPENDFROM command appends data from an existing .DBF data file to the current data file. All variables with matching names and types are appended to the current file. If variable length differs, data in the destination data file is either truncated or padded with spaces. Deleted records in the source file are not appended to the target file.

Example:

```
OPEN C:\SIMSTAT\DATA\ANNUAL.DBF;  
APPENDFROM C:\SIMSTAT\DATA\DECEMBER.DBF;
```

BEEP

Syntax:

```
BEEP;
```

Description:

Generates a beep signal through the computer's speaker.

BINOMIAL

Syntax:

```
BINOMIAL varlist BY varlist [/options];
```

Description:

The BINOMIAL test allows you to assess whether the observed number of cases in a dichotomous variable is the same as that expected from a specified binomial distribution. The observations can be divided either below or above the mean, the median or a user-specified cutoff value. Alternatively, the analysis can also be restricted to cases equal to two specified values. The user can also specify the test proportion.

Options:

VALUE (<i>real</i> [, <i>real</i>])	Cutoff value
MEAN	Cutoff value / values of X
MEDIAN	Mean
	Median
PROPORTION (<i>real</i>)	Test proportion (from 0 to 1)
PANEL	Displays the dialog box

BOOTSTRAP1**Syntax:**

```
BOOTSTRAP1 varlist [/options];
```

Description:

BOOTSTRAP1 performs bootstrap simulation to estimate the distribution of descriptive statistics in a population (e.g., mean, median, variance). The program draws a specified number of observations from the sample and computes the estimator for the subsample. This procedure is performed many times (10 to 30,000 times). The options allow you to display information about the estimator distribution including descriptive statistics, frequency table, percentile table and histogram of the estimator distribution. The program also computes nonparametric and bias-corrected bootstrap confidence intervals. If no sample size is specified (option SIZE), the bootstrap sample size is automatically adjusted to the size of the original sample.

Options:

SIZE= <i>integer</i>	Size of each sample
SAMPLING= <i>integer</i>	Number of samples
SEED= <i>integer</i>	Initial seed value
	Choice of estimator
MEAN	Mean
VARIANCE	Variance
STDDEV	Standard deviation
STDERR	Standard error
MEDIAN	Median
KURTOSIS	Kurtosis
SKEWNESS	Skewness
DESC	Descriptive statistics
INTERVAL= <i>real</i>	Confidence intervals
PTILES= <i>integer</i>	Percentile table
HISTOGRAM	Histogram
NBAR= <i>integer</i>	Nb of bars/intervals
NORMAL	Normal curve
PANEL	Displays the dialog box

BOOTSTRAP2**Syntax:**

```
BOOTSTRAP2 varlist BY varlist [/options];
```

Description:

The BOOTSTRAP2 command performs bootstrap resampling to estimate the distribution of various estimators in a given population (e.g., correlation, Tau-b, Cohen's Kappa). The program draws a specified number of pairs of observations from the sample and computes the estimator for the subsample. This procedure is performed many times (10 to 30,000 times). The options allow you to display information about the estimator distribution including descriptive statistics, frequency table, percentile table and histogram of the estimator distribution. The program can also compute nonparametric and bias corrected bootstrap confidence intervals, and the power rate of some estimators for 3 to 4 alpha levels. If no sample size is specified (option SIZE), the bootstrap sample size is automatically adjusted to the size of the original sample.

Options:

SIZE= <i>integer</i>	Size of each sample
SAMPLING= <i>integer</i>	Number of samples
SEED= <i>integer</i>	Initial seed value
	Choice of statistics
TAU-A	Kendall's Tau-A
TAU-B	Kendall's Tau-B
TAU-C	Kendall-Stuart's Tau-C
D-SYM	Somers' D symmetric
D-XDEP	Somers' D (X dependent)
D-YDEP	Somers' D (Y dependent)
GAMMA	Gamma
RHO	Spearman's Rho
R	Pearson's r
SLOPE	Regression slope
INTERCEPT	Regression intercept
S-T	Student's T
S-F	Student's F
M-W	Mann-Whitney
WILCOXON	Wilcoxon (W value)
SIGN	Sign test (Z value)
K-W	Kruskal-Wallis
MEDIAN	Median test (Z value)
AGREE	Percentage of agreement
KAPPA	Cohen's kappa
SCOTT	Scott's pi
NFREE	Free marginal (nominal scale)
KRBAR	Krippendorff's r bar
KR	Krippendorff's R
OFREE	Free marginal (ordinal scale)

DESC	Descriptive statistics
INTERVAL= <i>real</i>	Confidence interval
PTILES Percentile table	
HISTOGRAM	Histogram
NBAR= <i>integer</i>	Number of bars/intervals
NORMAL	Normal curve
POWER= <i>real</i>	Statistical power analysis
PANEL	Displays the dialog box

BOX...ENDBOX

Syntax:

```
BOX [Options]
:
ENDBOX;
```

Description:

The BOX command displays a window with textual information. By default the window is positioned in the middle of the screen. The TOP and LEFT options can also be used to specify the position of the window's upper left corner. The parameter for these two options is an integer value between 0 and 100 expressing a percentage of the screen height and width. The window stays on screen until the user presses **<Enter>** or clicks on the **OK** button. If the NOBUTTON option is used, the window stays on screen until a key is pressed. The DELAY option allows you to insert the minimum length of time the box should be displayed on screen before the user can proceed. The colors of the text in the box can also be altered using the COLOR option.

Options:

DELAY= <i>integer</i>	Length of delay (msec)
TOP= <i>integer</i>	Vertical position of the box's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the box's upper left corner (0 to 100)
<i>color</i>	Text color (see below)
NOBUTTON	Hide the OK button.
BEEP	Sounds a beep using computer's speaker.

Valid Colors:

Black Maroon Green Olive Navy Yellow Purple Teal
 Gray Silver Red Lime Blue Fuchsia Aqua White

Example:

```
BOX TOP=10 LEFT=10 COLOR=NAVY
The red line drawn in the scatterplot
is the regression line.
ENDBOX;
```


BREAKDOWN

Syntax:

```
BREAKDOWN varlist BY varlist [/options];
```

Description:

The BREAKDOWN command computes descriptive statistics for various sub-groups within the entire sample. Statistics are computed for each variable on the first list of variable, within groups defined by the values of the second list (grouping or independent variables). This command also allows you to obtain a multiple Box-&-Whisker plot that can be used to compare the distribution of the dependent variable among several sub-groups.

Options:

DETAIL	Detailed statistics
RANGE (<i>integer</i> , <i>integer</i>)	Range of X
BOXPLOT	Box-&-Whisker plot
PANEL	Display the dialog box

CALL

Syntax:

```
CALL filename;
```

Description:

This procedure executes another script file. After executing the external script file, the program continues at the statement following the CALL command. If no extension is provided, the program will successively look for an existing file name with an “.SCR” and an “.SCZ” extension. If no path information is provided, SIMSTAT will first look in the same directory as the calling script file and then in the default script directory.

Example:

```
CALL C:\DEMO\LESSON1.SCR;
```

CHART

Syntax:

```
CHART [options];
```

Description:

The CHART command allows you to modify various properties of the currently displayed chart, or navigate through the charts either to display another chart on the screen or modify some features of a chart currently not displayed.

Options:

FIRST	Move to the first chart
LAST	Move to the last chart
NEXT	Move to the next chart
PRIOR	Move to the previous chart
3D ON OFF	Add or remove 3D perspective
GRIDX ON OFF	Horizontal grid
GRIDY ON OFF	Vertical grid
GRIDY2 ON OFF	Second vertical grid
SCALEX (<i>value,value</i>)	Horizontal axis limits
SCALEY (<i>value,value</i>)	Vertical axis limits
INCX= <i>integer</i>	Increment value of the horizontal axis
INCY= <i>integer</i>	Increment value of the vertical axis
DECX= <i>integer</i>	Decimal places for values on the X axis
DEXY= <i>integer</i>	Decimal places for values on the Y axis
TITLE " <i>string</i> "	Title string
LABELX " <i>string</i> "	Horizontal axis string
LABELY " <i>string</i> "	Vertical axis string
LABELY2 " <i>string</i> "	Second vertical axis

CHI-SQUARE

Syntax:

```
CHISQUARE varlist BY varlist [/options];
```

Description:

The CHISQUARE command performs a one-sample chi-square test that allows to assess whether there is a difference between the observed number of cases in various categories and the expected frequencies in those same categories. The options allow you to restrict the test to specific values and to specify the expected frequencies.

Options:

VALUES (<i>real real ...</i>)	Expected values
FREQ (<i>real real ...</i>)	Expected frequencies
PANEL	Display the dialog box

CLUSTER**Syntax:**

```
CLUSTER varlist [/options];
```

Description:

This procedure performs hierarchical agglomerative cluster analysis of a distance or similarity matrix. Seven forms of clustering are presently available: the four average linkage procedures (unweighted pair group, unweighted centroid, weighted pair group, and weighted centroid [or median]); nearest and farthest linkage, and minimum variance. Requires MVSP v2.2.

Options:

LOG10	Select a transformation
LOGE	Log base 10
LOG2	Log base e
SQR	Log base 2
RATIO	Square root
TRANPOSE	Log ratio
	Transpose data
EUCLID	Select a coefficient (default: EUCLID)
SEUCLID	Euclidian distance
STEUCLID	Squared Euclidian distance
COSINE	Standardized Euclidian distance
MANHAT	Cosine theta distance
CANBER	Manhattan metric distance
CHORD	Canberra metric distance
CHISQR	Chord distance
AVERAGE	Chi-square distance
MEAN	Average distance
PEARSON	Mean character difference distance
SPEARMAN	Pearson product moment correlation
PERCENT	Spearman rank order correlation
GOWER	Percent similarity coefficient
SORENSEN	Gower general similarity coefficient
JACCARD	Sorensen's coefficient
MATCH	Jaccard's coefficient
YULE	Simple matching coefficient
NEI	Yule coefficient
	Nei & Lei's coefficient
UPAIR	Clustering method (default: UPAIR)
UCENTER	Unweighted pair group
WPAIR	Unweighted centroid
WCENTER	Weighted pair group
MINVAR	Weighted centroid
	Minimum variance

NEAR	Nearest linkage
FAR	Farthest linkage
TREEDESC	Create tree description file
TREEORDER	Create tree order file
RANDOMIZE	Randomize input order
CONSTRAIN	Constrained clustering
	Output of graphics
PLOT	Text dendrograms
GPLOT	Graphic dendrograms

COMPUTE

Syntax:

```
COMPUTE varname = expression;
```

Description:

The COMPUTE command allows the transformation of existing values of a variable or the computation of a new variable. SIMSTAT offers more than 50 operations and functions including numerical operators, trigonometric transformations (cos, sin, log, etc.), statistical functions (mean, minimum, maximum across variables or cases, etc.), and date and random number operations. (For more information on the available transformation functions available, see the “Computing values” section, page 50.)

Examples:

```
COMPUTE TOTSCORE = (DEP1+DEP2+DEP3+DEP4+LOG(DEP5))/5;
```

Conditional transformation can be performed by using successive FILTER and COMPUTE commands such as in the following examples:

```
FILTER AGE < 8;
COMPUTE TOTSCORE = AVG(DEP1..DEP10);
FILTER AGE >= 8;
COMPUTE TOTSCORE = AVG(DEP1..DEP15);
```

or

```
FILTER RELIGION = 1;
COMPUTE PREDFACT = 1.235;
FILTER .NOT. RELIGION = 1;
COMPUTE PREDFACT = 0.656;
```

CORANAL

Syntax:

```
CORANAL varlist [/options];
```

Description:

The correspondence analysis (or reciprocal averaging) procedure performs several varieties of correspondence analysis including detrended correspondence analysis (DCA). Correspondence analysis in general is well suited for working with count or presence/absence data, whereas PCA is geared more towards measurement data on a continuous scale (although PCA can also be performed on count and binary data. Requires MVSP v2.2.

Options:

LOG10	Select a transformation
LOGE	Log base 10
LOG2	Log base e
SQR	Log base 2
RATIO	Square root
	Logratio

TRANSPOSE	Transpose data
-----------	----------------

RECIPROCAL	Computation algorithm
JACOBI	Reciprocal averaging
	Cyclic Jacobi

PLOT	Output of graphics
GPLOT	Text scattergrams
	Graphic scattergrams

{the following options apply only to Reciprocal averaging}

DETRENDING	Detrending procedure
SEGMENTS= <i>integer</i>	Number of segments (default: 26)
CYCLES= <i>integer</i>	Number of cycles (default 4)
DOWNWEIGHT	Downweight rare species

{the following options apply only to Cyclic Jacobi}

KAISER	Minimum eigenvalue (default: none)
JOLLIFFE	Kaiser's rule
MINEIGEN= <i>real</i>	Jolliffe's rule
	Specify a minimum eigenvalue

ACCURACY= <i>real</i>	Accuracy of solution
-----------------------	----------------------

PERCENT	Select a weighting strategy
RARE	Adjust to percentage
COMMON	Weight on rare species
	Weight on common species

CORRELATION

Syntax:

```
CORRELATION varlist [BY varlist] [/options];
```

Description:

The CORRELATION command produces a matrix of Pearson correlation coefficients for all pairs of dependent-independent variables. If only one list of variables is specified, the procedure produces a square matrix where each variable is treated as dependent and independent. You can request either exact probabilities for the coefficients or a display of asterisks that indicate the probability levels attained. You can also tell the program to calculate probabilities using one- or two-tailed tests, to compute confidence intervals of specified width and to display cross-product deviation and covariance tables. A graphic scatterplot can also be obtained with or without regression lines.

Options:

EXACT	Exact probabilities / cases
1TAIL 2TAIL	Direction of the test
	Deletion of missing values
PAIRWISE	Pairwise
LISTWISE	Listwise
CCSS	Cross-product deviation & covariance
CI= <i>integer</i>	Confidence interval (width)
PANEL	Display the dialog box

CROSSTAB

Syntax:

```
CROSSTAB varlist BY varlist [/options];
```

Description:

The CROSSTAB command computes a contingency table for two variables where rows represent all the independent variable (x) values while the columns represent all the dependent (y) variable values. The options allow you to include various statistics in the table and obtain measures of association for nominal level (chi-square, phi, contingency coefficient) and ordinal level (gamma, tau-b, tau-c, Somers' d, etc.) variables. It also allows you to display a 3-d barchart of the two variables.

Options:

FREQ	Sort table by
DFREQ	Ascending frequency
VALUE	Descending frequency
DVALUE	Ascending value
	Descending value
TABLE	Display table
ROWPCT	Row percentages
COLPCT	Column percentages
TOTPCT	Total percentages
EXPECTED	Expected values
RESID	Chi-square residuals
SRESID	Standardized residuals
CHI2	Chi-square statistics
LRATIO	Likelihood ratio statistics
CONTINGENCY	Contingency coefficient
PHI Phi (2 x 2) or Cramer's V	
TAU-B	Kendall's tau-b
TAU-C	Kendall's tau-c
GAMMA	Goodman-Kruskal's gamma
SOMERS	Somers' d
RHO Spearman's Rho	
PEARSON	pearson's r
BARChart	2 dimension barchart
OVERLAP	Overlapping bars
STACKED	Stacked bars
100%	100% stacked bars
2D	2-D perspective
3D	3-D perspective
PANEL	Display the dialog box

DATA...ENDDATA**Syntax:**

```
DATA
:
:
ENDDATA;
```

Description:

The DATA command allows you to define temporary data to be analyzed. The first line following the DATA keyword must contain the variable names, while the remaining lines hold the data. SIMSTAT writes the embedded information to a temporary file named TEMP.DBF and automatically opens it for analysis. For more information on the proper format to use, see information on ASCII files format.

Example:

```
DATA
CYLINDERS, MPG, ACCEL
4 15.0 6.3
4 17.2 5.9
6 13.2 5.3
6 12.4 6.2
6 12.4 5.8
ENDDATA;
```

DEFINE or DIM**Syntax:**

```
DEFINE varname AS NUMERIC | STRING;
```

Description:

The DEFINE command allows you to explicitly declare a new memory variable and specify its type. Using a separate DEFINE statement for each variable, along with an explanation of the variable's purpose at the beginning of each subroutine and function, makes the script easier to debug and maintain.

Examples:

```
DEFINE $ANSWER AS NUMERIC;
DEFINE $USERNAME AS STRING;
```

DELAY**Syntax:**

```
DELAY (integer);
```

Description:

The DELAY command causes a delay in a program for a specified number of milliseconds.

Example:

```
Delay (5000);           {Suspends the program execution for 5 seconds}
```


DESC**Syntax:**

```
DESC varlist;
```

Description:

The DESC command displays the mean, standard deviation, minimum and maximum values of each specified variable. To obtain other descriptive statistics such as the variance, skewness, kurtosis, mode, median, etc., see the FREQUENCY command.

DIVERSITY**Syntax:**

```
DIVERSITY varlist [/options];;
```

Description:

This procedure computes three diversity indices commonly used in ecology, Simpson's, Shannon's, and Brillouin's. Requires MVSP v2.2.

Options:

TRANSPOSE	Transpose data
LOG10	Select a transformation (default: LOG10)
LOGE	Log base e
LOG2	Log base 2
SIMPSON	Select a coefficient (default: SIMPSON)
SHANNON	Simpson's diversity index
BRILLOUIN	Shannon's diversity index
	Brillouin's diversity index

FACTOR**Syntax:**

```
FACTOR varlist [/options];
```

Description:

The FACTOR command performs a principal component or an image covariance analysis with or without axis rotation. Requires EASY FACTOR ANALYSIS v3.0.

Options:

PCA	Type of analysis
IMAGE	Principal Component Image Covariance Analysis
NUMBER=integer	Maximum number of factors to retain
EIGEN=real	Minimum eigenvalue to retain
VARIMAX	Perform varimax rotation
SCREE	Display a scree plot in output
DESC	Display descriptive statistics
CORREL	Display correlation matrix
EXTRACTION	Display unrotated factor solution
ROTATION	Display rotated factor solution
WEIGHT	Display factor scoring weights
FSCORE	Display factor scores
SAVE	Save factor scores
PANEL	Display dialog box

FILTER**Syntax:**

```
FILTER [string];
```

Description:

The FILTER command allows you to temporarily select cases according to some logical condition. You can use this command to restrict your analysis to a subsample of cases or to temporarily exclude some subjects. These conditions are specified in a logical expression that may consist of a simple expression or include many expressions related by logical operators (AND, OR, XOR). Most xBase functions (see page 240) can be used with this command provided that the final expression can be evaluated as either true or false. The selection stays effective until the logical expression is changed or the selection deactivated. When used alone, this command deactivates the previous selection.

Examples:

```
FILTER (AGE > 10) .OR. (SEX = 1);  
FILTER; {deactivate the previous filter}
```

FREQUENCY

Syntax:

```
FREQUENCY varlist [/options];
```

Description:

The FREQUENCY command performs various frequency and descriptive analyses on specified variables. This command also provides a choice of graphs for either categorical or numeric variables.

Options:

TABLE NOTABLE	Display frequency table Sort table by
FREQ	Ascending frequency
DFREQ	Descending frequency
VALUE	Ascending value
DVALUE	Descending value
DESC	Descriptive statistics
CI= <i>integer</i>	Confidence interval width
PTILES= <i>integer</i>	Percentile table (number of categories)
BARChart	Bar chart
PIE	Pie chart
PARETO	Pareto chart
BOXPLOT	Box-&-whisker plot
CUMUL	Cumulative frequency plot
PLOT	Normal probability plot
HISTOGRAM	Histogram
NBAR= <i>integer</i>	Number of bars/intervals
NORMAL	Normal curve
MAXNOMINAL	Maximum number of values for table and nominal level charts
PANEL	Display the dialog box

FRIEDMAN

Syntax:

```
FRIEDMAN varlist;
```

Description:

The FRIEDMAN test is a procedure for testing whether two or more related samples have been drawn from the same population. The output displays the mean rank of each variable, the number of cases, chi-square, degree of freedom and probability value.

FULL

Syntax:

```
FULL [/options] command [/options];
```

Description:

The FULL command allows you to perform various statistical analyses such as frequency analysis, crosstabulation or multiple regression on successive bootstrap samples. The first set of options allows you set the sample size and the number of samples to be drawn from the original sample. A second set allows you to specify which analysis to perform and set the options normally available when calling the chosen statistical procedure. This second set of options allows to control how the analysis is to be performed and what statistics are to be printed. If no sample size is specified (option SIZE), the bootstrap sample size is automatically adjusted to the size of the original sample.

Options:

SIZE= <i>integer</i>	Size of each sample
SAMPLING= <i>integer</i>	Number of samples
SEED= <i>integer</i>	Initial seed value
PANEL	Display the dialog box

GLMANOVA

Syntax:

```
GLMANOVA varlist BY varlist [/ options];
```

Description:

The GLMANOVA command provides a General Linear Model implementation of analysis of variance and covariance. It can handles balanced and unbalanced ANOVA designs and support models with categorical and/or quantitative variables. The procedure can also be used to perform standard multiple regression problems that involve interaction terms. Various adjustment methods for unequal cell size are provided including a hierarchical strategy where the user can set the order of entry of each variable in the model. The various options allow you to display standard tables as well as various outputs usually found in ANOVA/ ANCOVA or multiple regression analyses.

Options:

COVAR (<i>varlist</i>)	Quantitative variables
INTERACTION (<i>var*var*...[var*var*...] ...</i>)	
STEP	Show statistics at each step
MULTREG	Multiple regression statistics
EQUATION	Regression equation
CI= <i>integer</i>	Confidence interval
MEAN	Adjusted means
CHANGE	Test of changes in R square
C PLOT	Residuals caseplot
OUTLIERS= <i>real</i>	Outliers criterion (s.d.)
DURBIN	Durbin-Watson statistic
PLOT	Residuals scatterplot
P PLOT	Residuals normality plot
SAVE	Save predicted and residuals
	Adjustment method
REGRESSION	Regression approach
NONEXP	Nonexperimental
HIERARCHICAL	Hierarchical
ORDER (<i>varlist</i>) (<i>varlist</i>)	Order of entry (hierarchical)
[... > (<i>varlist</i>)]	(All variables enumerated after the > character are entered with the interactions)
PANEL	Display the dialog box

GOTO**Syntax:**

```
GOTO label;
```

Description:

This command branches to another part of the program and continues processing the commands at that point. The line that the program is to switch to is marked with a label preceded by a colon (:).

Example:

```
GOTO JanuaryStat;

:JanuaryStat;
```

GOSUB

Syntax:

```
GOSUB label;
```

Description:

This command branches to another part of the program and continues processing the commands at that point until the RETURN keyword is encountered. The program then continues execution at the statement following the GOSUB command.

IF..THEN...ELSE

Syntax:

```
IF expression [THEN] command [ELSE command];
```

Description:

The IF command allows you to specify a condition that must be met for a command to be carried out. If the condition is true, the first command is executed. If the condition is false, the command appearing after the ELSE keyword is executed. If the keyword ELSE is not used, nothing is performed and the execution continues with the next statement. When used with the GOTO command, this command can increase the flexibility of the program by allowing SIMSTAT to switch to different parts of the program or perform specific actions under certain conditions.

Valid expressions:

A valid expression consists of an existing memory or a data file variable name, and a valid relational operator followed by a string or numeric expression. This numeric expression may consist of a numeric constant, another numeric variable, or an equation with arithmetic operations and mathematical functions (see page 176 for a list of available functions).

```
$NUMERIC =, <, >, <>, <=, => numeric expression
```

```
$STRING =, <> "string"
```

Examples:

```
IF $ANSWER1 >= 10 THEN GOSUB AgeError ELSE DESC all;  
IF $NAME = "John" THEN GOTO MonthlyAnalysis;  
IF DB.SALARY > TRUNC($MAXS * 1.1) THEN LET DB.SALARY = $MAXS*1.1;
```

IMPORT

Syntax:

```
IMPORT filename;
```

Description:

This command reads a data file produced by other applications and creates a new data file that may be used by SIMSTAT. The extension of the supplied filename is used to determine which kind of file is to be imported.

*.DB	Paradox data file (v3.0 - v5.0)
*.WK?	Lotus 123 (v1.0 - v5.0)
*.XLS	Excel (v2.1 - v5.0)
*.WB? or *.WQ?	Quattro Pro (v1.0 - v6.0)
*.WR?	Symphony (v1.0 - v1.1)
*.SIM	SIMSTAT for DOS (v1.0 - v3.5)
*.SYS	SPSS/PC+ file
*.SAV	SPSS for Windows file
*.CSV	Comma separated values (ASCII)
*.TAB	Tab separated values (ASCII)

Examples:

```
IMPORT C:\LOTUS\DATA\SURVEY.WK3;
IMPORT C:\SPSSWIN\DATA\IMPACT.SAV;
```

INPUT

Syntax:

```
INPUT $varname [AS type] "expression" [options];
```

Description:

The INPUT command allows you to get a value from the user and store the result in a user defined string or numeric variable. When a value is already stored in the specified variable, it will be presented as the default value to the user. Use the CLEAR option to erase this value. The LEN option can also be used to increase or decrease the maximum length of input. By default the dialog box is positioned in the center of the screen. However, the LEFT and TOP options can be used to specify the position of the box's upper left corner. The parameter for these two options is an integer value between 0 and 100 expressing a percentage of the screen height or width. If the variable specified is a numeric variable, you can restrict the valid range of responses by using the MIN and/or the MAX options. By

default, the number of decimals displayed is set to 0 and user's input is restricted to integer values. The DEC option can be used to alter the number of decimal places to display.

Options:

color	Text color (see below)
LEFT= <i>integer</i>	Vertical position of the box's upper left corner (0 to 100)
TOP= <i>integer</i>	Horizontal position of the box's upper left corner (0 to 100)
CLEAR	Clears previous value
LEN= <i>integer</i>	Maximum number of characters
DEC= <i>integer</i>	Number of decimal places displayed
MIN= <i>real</i>	Minimum value
MAX= <i>real</i>	Maximum value

Examples:

```
INPUT $NAME "What is your name? " LEN=50 CLEAR;
INPUT DB.AGE "How old are you? " MIN=5 MAX=25;
```

Valid Colors:

Black	Maroon	Green	Olive	Navy	Yellow	Purple	Teal
Gray	Silver	Red	Lime	Blue	Fuchsia	Aqua	White

INTERRATERS

Syntax:

```
INTERRATERS varlist BY varlist [/options];
```

Description:

The INTERRATERS command produces an inter-rater agreement table that consists of a square table where rows and columns contain the same categories used in both variables.

Options:

FREQ	Sort table by...
DFREQ	Ascending frequency
VALUE	Descending frequency
DVALUE	Ascending value
	Descending value
TABLE	Display table
ROWPCT	Row percentages
COLPCT	Column percentages
TOTPCT	Total percentages
EXPECTED	Expected values
RESID	Chi-square residuals
SRESID	Standardized residuals

PCTAGREE	Percentage of agreement
KAPPA	Cohen's kappa
SCOTT	Scott's pi
NFREE	Free marginal nominal agreement
KRBAR	Krippendorff's r-bar
KR	Krippendorff's R
OFREE	Free marginal ordinal agreement
BARChart	2 dimension barchart
OVERLAP	Overlapping bars
STACKED	Stacked bars
100%	100% stacked bars
2D	2-D perspective
3D	3-D perspective
PANEL	Display the dialog box

ITEM

Syntax:

```
ITEM varlist [/options];
```

Description:

The ITEM command performs a classical item analysis for multiple choice item questionnaires. Requires STATITEM v1.0.

Options:

KEYFILE	Specify that key responses are stored in a key file.
LISTWISE	Eliminate cases with missing values.
ISTAT	Display item statistics.
ALTERNATIVE	Display statistics for alternate items.
HILOW= <i>integer</i>	Set the discrimination index to a percentage between 10 and 50%.
IRATES	Display item-total response rates.
ICC	Display item characteristic curves.
SMOOTHED	Apply smoothing to item characteristic curves.
BREAKDOWN= <i>integer</i>	Display an endorsement breakdown table for up to 10 different groups.
TSTAT	Display detailed statistics of total scores.
TFREQ	Display a frequency table of total scores.
SAVE	Save the total scores in a data file.
PANEL	Display the dialog box.

KRUSKAL

Syntax:

```
KRUSKAL varlist BY varlist [/options];
```

Description:

The KRUSKALL command performs a Kruskal-Wallis one-way analysis of variance by ranks is a procedure for testing whether k groups have been drawn from the same population. This test is a nonparametric version of the one-way analysis of variance. The output displays the number of valid cases, the mean rank of the variable in each group, chi-square and its probability with a correction for ties.

Options:

```
VALUES (integer, integer) Range of X
PANEL Show dialog box
```

KS1

Syntax:

```
KS1 varlist [/options];
```

Description:

The KS1 command (Kolmogorov-Smirnov one-sample test) compares the distribution of each variable against a standard normal distribution or a uniform distribution. It tests whether the sample data can reasonably be thought to have come from a population having this theoretical distribution.

Options:

```
NORMAL Type of distribution
| UNIFORM Normal distribution
Uniform distribution

VALUES (real, real) Mean and standard deviation of a normal
distribution or minimum and maximum of a
uniform distribution

PANEL Display the dialog box
```

KS2**Syntax:**

```
KS2 varlist BY varlist [/options];
```

Description:

The KS2 command (Kolmogorov-Smirnov two-sample test) evaluates whether a dependent variable (Y) has the same distribution in two independent samples as defined by a grouping variable (X). This test is sensitive to differences in the shape, location, and scale of the two sample distributions.

Options:

VALUES (<i>integer, integer</i>)	Values of independent variable
PANEL	Display the dialog box

LET**Syntax:**

```
LET $variable = $variable, value or expression
```

Description:

The LET command assigns the value of the expression on the right side of the assignment operator (=) to the variable on the left side of the operator. The assignment statement must do the following:

- Identify the variable that receives the value.
- Use the assignment operator (=) to separate the variable from the expression.
- End with the expression that determines the variable's value.

Examples:

```
LET $AGE = 12;
LET $CLASSNAME = "STATISTICS 101";
```

You can also use a variable on both sides of the first assignment statement that uses it. For example, the following statement increases the value of the variable \$COUNT by one.

```
LET $COUNT = $COUNT + 1;
```

In addition to basic arithmetic operations, you can use various functions including mathematical procedures, trigonometric functions and random number functions on the right side of the expression (see page 176 for a list of available functions).

You may also use parentheses to control the evaluation sequence such as in:

```
LET $COUNT = ($LAST + 2) * 1.1415926;
LET $NEWAGE = SQRT(DB.AGE + 2 * NORMAL(2));
```

LIST

Syntax:

```
LIST varlist [/ N=integer];
```

Description:

The LIST command displays a listing of the values of the specified variables. The N option can be used to restrict the listing to the first N cases in the file.

LOGISTIC

Syntax:

```
LOGISTIC varlist [/options];
```

Description:

The LOGISTIC command performs a logistic multiple regression. Requires: LOGISTIC v3.11ef.

Options:

VALUES (<i>integer, integer</i>)	Values for success/failure
NOCONSTANT	Exclude the constant
ITER	Show iterations
CTABLE	Classification table
LRATIO	Likelihood ratio statistics
CI= <i>real</i>	Confidence interval width
TOLERANCE= <i>real</i>	Tolerance level
INTERACTION <i>var*var</i>	Interaction [*var*var*...]
PANEL	Display dialog box

MANN**Syntax:**

```
MANN varlist BY varlist [/options];
```

Description:

The MANN command produces a Mann-Whitney U test that evaluates the hypothesis that two independent samples have the same distribution. The Mann-Whitney U is the nonparametric version of the t-test for independent samples. This test is performed on the dependent variable divided into two groups as defined by values of the independent (grouping) variable. The probability test performed can be either one- or two-tailed.

Options:

VALUES (<i>integer, integer</i>)	Values of independent variable
1TAIL 2TAIL	Direction of test
PANEL	Display the dialog box

McNEMAR**Syntax:**

```
McNEMAR varlist BY varlist [/options];
```

Description:

The McNEMAR test is a procedure applied to a pair of correlated dichotomous variables to test whether there is a significant difference in proportions of subjects that change from one category to another. A binomial test is used to compute the significance level when the number of changes is less than 10. Otherwise, a chi-square statistic with the Yates correction for continuity is used.

Options:

VALUES (<i>integer, integer</i>)	Values of independent variable
1TAIL 2TAIL	Direction of the test
PANEL	Display the dialog box

MEDIAN

Syntax:

```
MEDIAN varlist BY varlist [/options];
```

Description:

The MEDIAN test is a procedure for testing whether two or more independent groups differ in central tendencies. It tests the likelihood that those groups were drawn from populations with the same median. The output displays the number of cases greater than, less than, and equal to the median for each category of the grouping variable. Also displayed are the median, chi-square, degree of freedom and probability value.

Options:

EXTENDED	Extended median test
VALUES (int, int)	Values or range of X
PANEL	Display the dialog box

MENU...ENDMENU

Syntax:

```
MENU [options]  
:  
ENDMENU;
```

Description:

The MENU command allows you to define a bouncing bar menu. Every line between the MENU and ENDMENU commands will become a menu item. The maximum number of items is 30. An optional '&' character can be inserted in the command file to specify an accelerator key that, when pressed, will select this item. A single line command can be associated with a menu item by putting a command between this menu item and the associated command. A menu separator can also be added by putting a "-" character on a separate line.

By default, the menu is displayed in the middle of the screen. The TOP and LEFT options can also be used to specify the position of the window's upper left corner. The parameter for these two options is an integer value between 0 and 100 expressing a percentage of the screen height and width.

Options:

<code>TOP=integer</code>	Vertical position of the box's upper left corner (0 to 100)
<code>TOP=integer</code>	Horizontal position of the box's upper left corner (0 to 100)

Example:

```
MENU LEFT=30 TOP=20
&Open,OPEN C:\DATAFILE\DATAFILE
&Save,GOTO SAVEDATA
Save &As,GOTO SAVEAS
-
&Quit,QUIT
ENDMENU;
```

MOSES**Syntax:**

```
MOSES varlist BY varlist [/options];
```

Description:

The MOSES test of extreme reactions tests whether the range of an ordinal variable is the same in a control group as in a comparison group, as defined by a grouping variable. The output includes counts for both groups, number of outliers removed, the span of the control group before and after outliers are removed, and one-tailed probability of the span with and without outliers. By default, 5% of the cases are trimmed from each end of the range of the control group to remove outliers.

Options:

<code>VALUES (int, int)</code>	Values of independent variable
<code>OUTLIERS=integer</code>	Number of outliers to remove
<code>PANEL</code>	Display the dialog box

MRESPONSE**Syntax:**

```
MRESPONSE [MULTX, MULTY]
FREQUENCY | CROSSTAB | INTERRATERS [/options]
```

Description:

The MRESPONSE (multiple responses) command allows you to obtain frequency analyses and crosstabulation analyses on variables which can legitimately have more than one response. These multiple responses are stored in as many variables as necessary. Choosing all these variables as dependent (X) or independent (Y) variables allows you to gather all these responses and treat them as if they were stored in a single variable.

Options:

```

MULTX                Treat X as multiple responses
MULTY                Treat Y as multiple responses

PANEL                Display the dialog box

FREQ varlist [/options];
CROSSTAB varlist BY varlist [/options];

```

Example:

```

MRESPONSES MULTY
CROSSTAB INCOME1 INCOME2 BY SEX /TABLE DFREQ;

```

MULTREG**Syntax:**

```
MULTREG varlist BY varlist [/options];
```

Description:

The MULTREG command allows you to perform multiple regression analysis to predict a dependent variable from many independent variables. SIMSTAT provides various regression methods including standard regression, hierarchical entry, forward selection, backward selection, stepwise selection. The options also allow you to display a wide range of statistics and perform various tests on residual values.

Options:

HIERARCHICAL	Type of regression
FORWARD	Hierarchical entry
BACKWARD	Forward selection
STEPWISE	Backward elimination
STANDARD	Stepwise selection
	Enter all variables
PIN= <i>real</i>	P value to enter
POUT= <i>real</i>	P value to remove
TOLERANCE= <i>real</i>	Minimum tolerance value
STEP	Show each step
ANOVA	ANOVA table
CHANGE	Test of changes in R-square
HISTORY	History of changes in R
SUMMARY	Summary ANOVA table
EQUATION	Variables in the equation
OUT	Variables not in the equation
CI= <i>integer</i>	Width of confidence interval
CPLOT	Casewise plot of residuals
OUTLIERS= <i>real</i>	Outliers criterion (s.d.)
DURBIN	Durbin-Watson statistic
RPLOT	Residuals scatterplot
PPLOT	Normal plot of residuals
SAVE	Save predicted values and residuals
PANEL	Display the dialog box

NOTE**Syntax:**

```
NOTE "expression" [options];
```

Description:

The NOTE command displays a single line string on the background screen. By default, the note line is displayed horizontally centered and at the bottom of the screen with a font size of 10 points. The TOP and LEFT options can also be used to specify the position of the text's upper left corner. The parameter for these two options is an integer value between 0 and 100 expressing a percentage of the screen height and width. Other options allow you to control the size, style and color of the displayed text.

Options:

TOP= <i>integer</i>	Vertical position of the text's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the text's upper left corner (0 to 100)
SIZE= <i>integer</i>	Size of the font (between 6 and 100)
COLOR= <i>color</i>	Color of the font (see below)
ITALIC	Displays the string in italic characters
BOLD	Displays the string in bold characters
HIDE	Hides the string

Examples:

```
NOTE "Monthly analysis" SIZE=10 COLOR=YELLOW BOLD;
NOTE HIDE;
```

Valid Colors:

Black	Maroon	Green	Olive	Navy	Yellow	Purple	Teal
Gray	Silver	Red	Lime	Blue	Fuchsia	Aqua	White

NPAR

Syntax:

```
NPAR varlist BY varlist [/options];
```

Description:

The NPAR command displays a matrix for various measures of association and concordance between two variables. The options provide counts and exact probabilities of each coefficient or asterisks that show the probability level reached. This probability test can be either one- or two-tailed.

Options:

TAU-A	Type of statistics
TAU-B	Kendall's Tau-a
TAU-C	Kendall's Tau-b
D-SYM	Kendall-Stuart's Tau-C
D-XDEP	Somers' D (symmetric)
D-YDEP	Somers' D (X dependent)
GAMMA	Somers' D (Y dependent)
RHO	Gamma
R	Spearman's Rho
	Pearson's R
EXACT	Exact probabilities
1TAIL 2TAIL	Direction of test
PAIRWISE	Deletion of missing values
LISTWISE	Pairwise
	Listwise
PANEL	Display the dialog box

ONEWAY

Syntax:

```
ONEWAY varlist BY varlist [/options];
```

Description:

The ONEWAY command performs a one-way analysis of variance for all dependent variables on groups defined by each categorical (numeric) independent variable. It allows testing whether the means of the groups (2 or more) are not all equals to each other. ONEWAY provides one-way variance analysis and a standard table including between and within groups sums of squares, mean squares, and degrees of freedom, F-ratio and its associated probability. You can also obtain for each group, descriptive statistics including count, mean, standard deviation, standard error and a user-specified confidence interval for the mean. Various measures of effect size and post hoc comparisons can also be obtained

as well as various graphs such as a barchart representing the mean of each group, an error bar diagram or a deviation chart where each bar represents either the standard deviation, the standard error or a user specified confidence interval.

Options:

RANGE (<i>integer, integer</i>)	Minimum and maximum values
DESC	Descriptive statistics
CI= <i>integer</i>	Confidence interval width
	Post hoc test
LSD	Least significant difference
N-K	Newman-Keuls
TUKEY	Tukey's H.S.D
SCHEFFE	Scheffé test
ERRORCHART	Error bar chart
BARCHART	Displays mean bars
UPPER	Displays only upper portion of error bars
CIBAR	Confidence interval
SE	Standard error
SD	Standard deviation
LINK	Link means
DEVCHART	Deviation chart
PANEL	Display the dialog box

OPEN

Syntax:

```
OPEN filename;
```

Description:

The OPEN command reads a file. The extension of the file is used to determine what kind of file should be opened. When the file name includes a .DBF extension, SIMSTAT closes the existing data file and opens this new data file. When an .SNB extension is used, the program clears the current notebook and loads the specified notebook file. If the file extension is .CHX then SIMSTAT clears the content of the Chart window and loads the specified chart file.

Examples:

```
OPEN C:\DATA\CARS.DBF;           (open the cars.dbf data file)
OPEN C:\OUTPUT\REPORT.SNB;     (open the report.snb notebook file)
```

PCA

Syntax:

```
PCA varlist [/options];
```

Description:

This procedure performs a R-mode principal component analysis. The component loadings are scaled to unity, so that the sum of squares of an eigenvector equals 1, and the component scores are scaled so that the sum of squares equals the eigenvalue. Q-mode PCA will generally have the opposite scaling. Requires MVSP v2.2.

Options:

	Select a transformation
LOG10	Log base 10
LOGE	Log base e
LOG2	Log base 2
Sqrt	Square root
RATIO	Logratio
TRANSPOSE	Transpose data
CENTER	Centered data
STANDARDIZE	Standardize data
	Minimum eigenvalue (default = all)
KAISER	Kaiser's rule
JOLLIFFE	Jolliffe's rule
MINEIGEN= <i>real</i>	Specify a minimum eigenvalue
ACCURACY= <i>real</i>	Accuracy of solution
	Output of graphics
PLOT	Text scattergrams
GPLOT	Graphic scattergrams

PCO

Syntax:

```
PCO varlist [/options];
```

Description:

Principal coordinates analysis (PCO) is a generalized form of PCA. Whereas PCA implicitly uses either a covariance or correlation matrix, PCO allows you to input any matrix of metric values. PCO may be used with any of the distances calculated by MVSP except for the squared Euclidean distance. Of the similarity measures only Gower's is metric. PCO is calculated as a Q-mode eigenanalysis, therefore it only gives the eigenvectors, not scores. Note that a PCO of Euclidean distances will give the same results as a Q-mode PCA. Requires MVSP v2.2.

Options:

LOG10	Transformation
LOGE	Log base 10
LOG2	Log base e
SQRT	Log base 2
RATIO	Square root
	Logratio
TRANSPOSE	Transpose data
EUCLID	Select a coefficient (default: EUCLID)
STEUCLID	Euclidian distance
COSINE	Standardized Euclidian distance
MANHAT	Cosine theta distance
CANBER	Manhattan metric distance
CHORD	Canberra metric distance
CHISQR	Chord distance
AVERAGE	Chi-square distance
MEAN	Average distance
GOWER	Mean character difference distance
	Gower general similarity coefficient
KAISER	Minimum eigenvalue (default: none)
JOLLIFFE	Kaiser's rule
MINEIGEN=real	Jolliffe's rule
	Specify a minimum eigenvalue
ACCURACY=real	Accuracy of solution
PLOT	Output of graphics
GPLOT	Text scattergrams
	Graphic scattergrams.

PICTURE**Syntax:**

```
PICTURE NOx filename [options]
```

Description:

The PICTURE command allows you to display Windows Metafiles (.WMF), Windows bitmap (.BMP) or icon (.ICO) files. The filename extension is used to determine the graphic type of the file. Up to 5 different pictures can be displayed on the same screen. This command can be used with the CURRENTCHART option to display a bitmap copy of the currently active chart. When used with this option, the size of the displayed picture will be equal to the size of the chart window's content (see the WINDOW command for instruction on how to change the size of a chart window). By default, the pictures are displayed in the middle of the screen. The TOP and LEFT options can also be used to specify the position of the picture's upper left corner.

Options:

NO1..NO5	Picture number
LEFT= <i>integer</i>	Vertical position of the picture's upper left corner (0 to 100)
TOP= <i>integer</i>	Vertical position of the picture's upper left corner (0 to 100)
CURRENTCHART	Display the currently active chart
HIDE	Hide the picture.

Examples:

```
PICTURE NO1 'C:\SIMSTAT\SCRIPT\EARTH.BMP' TOP=10 LEFT=20;
PICTURE NO2 CURRENTCHART TOP=30 LEFT=60;
PICTURE NO1 HIDE;
```

PLAY**Syntax:**

```
PLAY filename [options];
```

Description:

The PLAY command allows you to play multimedia files such as .WAV sound files, .MDI music files or .AVI movie files. If no directory is provided, the program successively looks for the file in the active directory and in the directory where the current script file is located. When playing an .AVI movie file, the viewer is positioned in the middle of the screen. The TOP and LEFT options can also be used to specify the position of the window's upper left corner. The parameter for these two options is an integer value between 0 and 100 expressing a percentage of the screen height and width.

Options:

TOP= <i>integer</i>	Vertical position of the AVI movie screen's upper left corner (0 and 100)
LEFT= <i>integer</i>	Horizontal position of the AVI movie screen's upper left corner (0 and 100)

Examples:

```
PLAY C:\WINDOWS\TADA.WAV;
PLAY SONATA.MDI;
PLAY INTRO.AVI TOP=10 LEFT=20;
```

PRINT**Syntax:**

```
PRINT windowtype [options];
```

Description:

The PRINT command can be used to send the contents of a window to the printer. By default, all contents are printed. To restrict the printing to some pages of a notebook or to a specific number of charts, use the FROM and TO options to specify the lower and upper limits.

Options:

DATA	Window types
NOTEBOOK	Data spreadsheet window
SCRIPT	Notebook/Statistical results window
CHART	Script/log window
	Charts window
FROM= <i>integer</i>	Start printing at page/chart
TO= <i>integer</i>	End printing at page/chart

Examples:

```
PRINT NOTEBOOK;
PRINT CHART FROM 3 TO 5;
```

QBOX**Syntax:**

```
QBOX "string" [options];
```

Description:

The QBOX command allows you to display a single line message or question to the screen. By default the window is positioned in the middle of the screen. The LEFT and TOP options can also be used to specify the position of the window. The window stays on screen until a key is pressed. The DELAY option allows you to insert a minimum delay between the display of the box and the input of a valid key. The colors of the text appearing in the box can be altered by specifying a color as an option. The default colors can also be changed by using the SET COLOR command.

Options:

Color	Text color
DELAY= <i>integer</i>	Length of delay (msec)
TOP= <i>integer</i>	Vertical position of the box's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the box's upper left corner (0 to 100)
BEEP	Sounds a beep using computer's speaker
NOBUTTON	Hide the OK button

Valid Colors:

Black	Maroon	Green	Olive	Navy	Yellow	Purple	Teal
Gray	Silver	Red	Lime	Blue	Fuchsia	Aqua	White

QUESTION...ANSWERS...ENDBOX

Syntax:

```
QUESTION varname [/options];
:
:
ANSWERS
:
:
ENDBOX;
```

Description:

This command displays a dialog box with a multiple choice question. The lines between the QUESTION and ANSWERS keywords contain the text of the question while the lines between the ANSWERS and ENDBOX keywords are used to type the available answers. Each answer should be entered on a separate line. An optional '&' character can be inserted in the line to specify an accelerator key that, when pressed, will select this item. The maximum number of items for each question is 30. The answer provided by the user is stored as a number in a variable specified on the first line of the command. When a data file variable is specified, the answer is automatically stored in the current record of the data file. If a numeric value is already stored in variable, it is used as the default answer. The CLEAR option can be used to automatically erase this value before the dialog box is displayed. If an answer is required and should not be skipped, use the NOSKIP option to prevent the user from exiting the dialog box without providing a valid answer.

Options:

color	Color of the text
TOP= <i>integer</i>	Vertical position of the dialog box's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the dialog box's upper left corner (0 to 100)
BEEP	Sounds a beep using computer's speaker
NOBUTTON	Hide the OK button
CLEAR	Clear the content of the variable
MOSKIP	Prevent exiting the dialog box without providing a valid answer

Example:

```

QUESTION $ANSWER1 BLUE NOSKIP
In the following expression:  P = 0.05
What does the probability value stands for?
ANSWERS
&A) The probability that the null hypothesis is true
&B) The probability that the null hypothesis is false
&C) The probability of the data given the null hypothesis
&D) The probability of the null hypothesis given the data
&E) None of the above
END;
```

QUIT**Syntax:**

```
QUIT
```

Description:

This command takes you out of the SIMSTAT program. All opened files are closed before exiting the program.

RANDOM**Syntax:**

```
RANDOM varlist BY varlist [/options];
```

Description:

The RANDOM command is similar to the BOOTSTRAP2 command but simulates the null assumption that there is no difference or relation in the population. The program draws from the sample and for each variable a specified number of random observations with replacement and computes the estimator for the subsample. The procedure is performed many times (10 to 30,000 times). The options allow you to display information about the estimator distribution including descriptive statistics, frequency table, percentile table and histogram of the estimator distribution. The program also computes nonparametric and bias-corrected bootstrap confidence intervals and Type I error rate for up to 4 alpha levels. If no sample size is specified (option SIZE), the bootstrap sample size is automatically adjusted to the size of the original sample.

Options:

SIZE= <i>integer</i>	Size of each sample
SAMPLING= <i>integer</i>	Number of samples
SEED= <i>integer</i>	Initial seed value
	Choice of statistics
TAU-A	Kendall's Tau-A
TAU-B	Kendall's Tau-B
TAU-C	Kendall-Stuart's Tau-C
D-SYM	Somers' D symmetric
D-XDEP	Somers' D (X dependent)
D-YDEP	Somers' D (Y dependent)
GAMMA	Gamma
RHO	Spearman's Rho
R	Pearson's r
SLOPE	Regression slope
INTERCEPT	Regression intercept
S-T	Student's T
S-F	Student's F
M-W	Mann-Whitney
WILCOXON	Wilcoxon (W value)
SIGN	Sign test (Z value)
K-W	Kruskal-Wallis
MEDIAN	Median test (Z value)
AGREE	Percentage of agreement
KAPPA	Cohen's kappa
SCOTT	Scott's pi
NFREE	Free marginal (nominal scale)
KRBAR	Krippendorff's r bar
KR	Krippendorff's R
OFREE	Free marginal (ordinal scale)
DESC	Descriptive statistics
CI= <i>real</i>	Confidence interval
PTILES	Percentile table
HISTOGRAM	Histogram
VERT HORIZ	Orientation
MIN= <i>real</i>	Minimum value
INC= <i>real</i>	Increment value
NBAR= <i>integer</i>	Nb of bars/intervals
NORMAL	Normal curve
ERROR= <i>real</i>	Type I error rate
PANEL	Displays the dialog box

RANK**Syntax:**

```
RANK varname;
```

Description:

Replaces the values of the current variable by their rank. If ties occur, the mean rank of the tied values is used. Missing values are excluded.

Examples:

```
RANK FINALNOTE;
```

To store the rank in another variable, use this command in combination with the **COMPUTE** command as in the following example:

```
COMPUTE FINALRANK = FINALNOTE;
RANK FINALRANK;
```

RECODE**Syntax:**

```
RECODE varname = (value[,value..] = value) [...];
```

Description:

The **RECODE** command provides an easy way to make multiple changes to the values of numeric variables, or to collapse values of a continuous variable into categories. The recode expression can consist of several transformations enclosed in parentheses, each including one or more values, or a value range, an equal sign, and the new value. Each value on the left of the equal sign is recoded into the value on the right. The recoding proceeds from left to right and stops after a transformation occurs. The **SYSMIS** and **MISSING** keywords can be used to represent missing values, while the **ELSE** represents all non specified values.

Examples:

```
RECODE SESTATUS = (1,2 = 1) (3,4,5 = 2) (6..10 = 3) (ELSE = MISSING);
```

To store the new codes into another variable, use this command in combination with the **COMPUTE** command as in the following example:

```
COMPUTE SESTATCAT = SESTATUT;
RECODE SESTATCAT = (1,2 = 1) (3,4,5 = 2) (6..10 = 3) (ELSE = 4);
```

RECORD

Syntax:

```
RECORD [option];
```

Description:

The RECORD command allows you to move the cursor within the currently opened data file to make a specific record active. You can also use this command to add a new record at the bottom of the data file.

Options:

FIRST	Move to the first record
LAST	Move to the last record
NEXT	Move to the next record
PRIOR	Move to the previous record
NEW	Create a new record

REGRESSION

Syntax:

```
REGRESSION varlist BY varlist [/options];
```

Description:

The REGRESSION command produces simple regression analysis for each pair of dependent-independent variables. SIMSTAT lets you choose among linear and 6 types of nonlinear regression. The output includes the Pearson product-moment correlations, the intercept and slope of the regression line, and an ANOVA table for the equation. Various options allow you to obtain a bivariate scatterplot, to select a one- or two-tailed test of probabilities and to request a standardized residuals caseplot, a scatterplot of predicted values by standardized residuals or a normal probability plot of residuals.

Options:

	Type of regression
LINEAR	Linear regression
QUADRATIC	Quadratic regression
CUBIC	Cubic regression
4TH	4th degree polynomial
5TH	5th degree polynomial
INV	Inverse regression
LOG	Logarithmic regression
EXP	Exponential regression

1TAIL 2TAIL	Direction of test
XYPLOT	Bivariate scatterplot
CI= <i>integer</i>	Confidence interval width
CPlot	Caseplot of residuals
OUTLIERS= <i>real</i>	Outliers criterion (s.d.)
DURBIN	Durbin-Watson statistic
RPlot	Residuals scatterplot
PPlot	Probability plot of residuals
SAVE	Save predicted value and residuals
PANEL	Display the dialog box

RELIABILITY

Syntax:

```
RELIABILITY varlist [BY varlist] [/options];
```

Description:

The RELIABILITY command provides a means to assess the quality of multiple-item additive scales through the computation of reliability statistics. The options available offer the possibility to obtain various item statistics, iter-item variance-covariance and correlation matrices, total scale and item-total statistics. It also allows you to verify the reliability of the scale through the use of a split-half method or by computing internal consistency measures. Each selected variable is considered as a single item of the scale. The first and second lists or variables can be used to specify the division of items in two different subscales to be used in a split-half method.

Options:

ITEM	Item statistics
CORR	Inter-item correlation matrix
COVAR	Variance/covariance matrix
TOTAL	Item-total statistics
SPLIT	Split-half reliability
ALPHA	Cronbach's Alpha
PANEL	Display the dialog box

RETURN

see GOSUB

RUN

Syntax:

```
RUN program parameters;
```

Description:

The RUN command runs another Windows or DOS program. If the program is not in the program directory or in the current script directory, you will need to specify the full path of the program.

Examples:

```
RUN SIMCALC.EXE;
RUN C:\WINDOWS\WRITE.EXE MYDOC.WRI;
```

RUNS

Syntax:

```
RUNS varlist BY varlist [/options];
```

Description:

The RUNS test is a procedure to test whether the ordered sequence in which observations were obtained is random. In order to be performed, such a test requires that all values be dichotomized into two categories. The options allow you to separate observations into two distinct groups using the mean, the median or a user-specified value as a cutoff point.

Options:

	Type of cutoff point
MEAN	Mean
MEDIAN	Median
VALUE	Value
VALUES (<i>integer</i> , <i>integer</i>)	Values of X
PANEL	Display the dialog box

SAVE

Syntax:

```
SAVE windowtype;
```

Description:

This command can be used to store the contents of the notebook, the chart or the script window on disk. If the contents were created during the current session and have not been saved on disk before, a dialog box will appear to allow you to specify the name and location of the new file.

Options:

NOTEBOOK	Window types
SCRIPT	Notebook/Statistical results window
CHART	Script/log window
	Charts window

Examples:

```
SAVE NOTEBOOK;
SAVE CHART;
```

SCED

Syntax:

```
SCED varlist BY varlist [/options];
```

Description:

The SCED (single case experimental design) command provides some basic tools to study the effect of an intervention on the behavior of a single subject. It involves the repeated objective measurement of the behavior of a single subject (dependent variable) over a long period of time interspersed with changes in the treatment condition (independent variable). The procedure will display a graph representing the evolution of the dependent variable (Y) at various phases defined by the independent (X) variable. The various options allow you to obtain statistics for each phase of the analysis as well as graphic tools that can be used as judgemental aids to identify the experimental effect of the intervention (smoothed data, split-middle trend, control bars).

Options:

BRIEF DETAIL	Descriptive statistics
CUMUL	Cumulative frequency
LOG	Log transformation
MEAN	Display mean lines
REGRESSION	Display least square regression lines
MEDIAN	Display split-middle trend lines
	Smoothing technique
MAVG (<i>int int ...</i>)	Moving average
RMED (<i>int int ...</i>)	Running median
RBAR	Control bars
PCT= <i>integer</i>	Width of interval
MIN= <i>integer</i>	Lower value
MAX= <i>integer</i>	Higher value
PANEL	Display the dialog box

SCREEN**Syntax:**

```
SCREEN color;
```

Description:

This command allows you to display a background screen on which textual information, pictures, movies, menus and dialog boxes will be displayed. The *color* option lets you specify the background color. To remove the screen background, use the HIDE option.

Option:

<i>color</i>	See below for valid background colors
HIDE	Remove the screen

Valid Colors:

Black	Maroon	Green	Olive	Navy	Yellow	Purple	Teal
Gray	Silver	Red	Lime	Blue	Fuchsia	Aqua	White

Examples:

```
SCREEN NAVY;
SCREEN HIDE;
```


SENSITIVITY

Syntax:

```
SENSITIVITY varlist BY varlist [/options];
```

Description:

The SENSITIVITY analysis allows one to assess the ability of a quantitative measure (X) to differentiate a dichotomous criterion condition (Y) and provide guidelines to choose an appropriate cutoff point. The program provides, for each value of the quantitative measure, the level of sensitivity (proportion of positive cases correctly diagnosed as true) and specificity (proportion of negative cases correctly diagnosed as false), and the percentage of false-positives and false-negatives. This command also allows you to obtain a ROC (receiver operating characteristic) curve and an Error rate graph.

Options:

VALUE=real	Criterion value
HIGH	Scale orientation (ascending)
SSTAT	Sensitivity statistics
ESTAT	Error rates statistics
ROC	Roc curve graph
ERROR	Error rates graph
PANEL	Display the dialog box

SET

Syntax:

```
SET [options];
```

Description:

The SET command can be used to change various global program options such as the display of the toolbar, the status bar or the help hints. It can also be used to modify options used when printing notebook pages or charts such as the page header or the number of charts per page.

The HEADER option is used to specify a line of text that will appear at the top of each printed analysis result. By default, this command alters the contents of the header printed at the center of the page. The sections of the header printed to the left or the right edge of the page remain unchanged. However, you can also modify those sections by using the **less than** (<) and the **greater than** (>) character. All text that appears to the left of the first < character will be printed on the left margin. Also, all text appearing to the right of the last > character will be printed flush right. Special codes can be inserted in the header to display the source data file (\$f), the time (\$t) and date (\$d) when the analysis was done, or the notebook page number (\$p).

The VARSIZE and DECIMALS parameters are used to specify the default physical size and number of decimal places of newly created variables. It is recommended to set those parameters before creating a new variable with the COMPUTE command.

The SET command also allows you to run a script in demonstration mode, by using the

```
SET DEMO=integer;
```

command where integer stands for the number of milliseconds between dialog boxes. In this mode, BOX and QBOX no longer stop until the user presses <ENTER> or clicks on the OK button but will be displayed only for a specified length of time. To disable the DEMO mode, set the number of milliseconds to zero.

Options:

TOOLBAR ON OFF	Turns the main window's tool bar on/off
STATUS ON OFF	Turns the main window's status bar on/off
HINTS ON OFF	Turns the display of help hints on/off
BEEP ON OFF	Beep on error
SIZE= <i>integer</i>	Default physical size of newly created variables
DECIMAL= <i>integer</i>	Default number of decimal places for newly created variables
HEADER= <i>string</i>	Changes the title line on printed output
CHARTSPERPAGE= <i>integer</i>	Number of charts per page (1, 2 or 4)
DEMO (<i>integer</i>)	Enables/disables demonstration mode (0 to disable or 1 to 30000 milliseconds)

SIGN

Syntax:

```
SIGN varlist BY varlist [/options];
```

Description:

The SIGN test procedure tests the hypothesis that two variables have the same distribution. This is assessed by comparison of the numbers of positive and negative differences between values of the two variables. The probability test performed can be either one- or two-tailed.

Options:

1TAIL 2TAIL	Direction of the test
PANEL	Display the dialog box

SORT**Syntax:**

```
SORT varname or expression;
```

Description:

This command allows you to arrange cases in numeric or alphabetic order. If only one variable is specified, the sorting will be done on this variable in ascending order. To sort the records in descending order or to specify a more complex sort involving several variables, you need to specify a sort expression. This expression can include almost any supported xBase function.

Examples:

```
SORT AGE;                Sort on AGE in ascending order
SORT DESCEND(AGE);      Sort on AGE in descending order.
SORT SEX*100+AGE;       Composite xBase expression that will
                        first sort the records on sex, and then
                        on age
```

STOP**Syntax:**

```
STOP;
```

Description:

Immediately stops the batch command and returns to the SIMSTAT user interface.

T-TEST**Syntax:**

```
T-TEST varlist BY varlist [/options];
```

Description:

The T-TEST command calculates either independent-sample t-tests (GROUP) or paired-sample t-tests (PAIRED) to decide whether two sample means are significantly different. The paired-sample (or correlated) t-test compares the means between each pair of dependent and independent variables. The independent t-test compares means on the dependent variable for two groups defined by values of the independent variable. SIMSTAT provides two distinct tests to take into account whether the two populations from which the samples are drawn have equal or unequal variances. You can also specify whether the null hypothesis should be evaluated using a one- or a two-tailed test.

Options:

GROUP PAIRED	Grouped or paired t-test
VALUE (<i>integer, integer</i>)	Values of X
CI= <i>integer</i>	Confidence interval width
1TAIL 2TAIL	Direction of the test
ERRORCHART	Error bar graph
BARCHART	Display mean bars
UPPER	Display upper error bar only (barchart)
CIBAR	Confidence interval
SE	Standard error
SD	Standard deviation
HISTOGRAM	Dual histogram
NBAR= <i>integer</i>	Number of bars/intervals
NORMAL	Normal curve
VERTICAL	Vertical bars
PANEL	Display dialog box

TIME-SERIES**Syntax:**

```
TIME-SERIES varlist [/options];
```

Description:

The TIME-SERIES command allows the examination of time series. Available options offer various transformations to remove trends or seasonal dependence in a series and provide a diagnostic for those transformations by displaying autocorrelation and partial autocorrelation function plots of the transformed series. This command also allows the application of two smoothing methods (i.e., moving average and running median) to identify trends in noisy time series data. Control bars representing the mean and the confidence limits can also be displayed over the series.

Options:

LOG	Log transformation
MEAN	Remove the mean
DIFF= <i>integer</i>	Number of difference operations
SEASON= <i>integer</i>	Length of seasonality
PLOT	Plot the series
ACF	Autocorrelation function
PACF	Partial autocorrelation function
LAG= <i>integer</i>	Number of lags for ACF and PACF
MAVG (<i>int int...</i>)	Type of smoothing Moving average
RMED (<i>int int...</i>)	Running median
RBAR	Control bars

PCT= <i>integer</i>	Width of interval
LOW= <i>integer</i>	Lower value
HIGH= <i>integer</i>	Higher value
PANEL	Display the dialog box

TITLE

Syntax:

```
TITLE "expression" [options];
```

Description:

The TITLE command displays a single line string on the background screen. By default, the title line is displayed horizontally centered and at the top of the screen with a font size of 24 points. The TOP and LEFT options can also be used to specify the position of the text's upper left corner. The parameter for these two options is an integer value between 0 to 100 expressing a percentage of the screen height and width. Other options allow you to control the size, style and color of the displayed text.

Options:

TOP= <i>integer</i>	Vertical position of the text's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the text's upper left corner (0 to 100)
SIZE= <i>integer</i>	Size of the font (between 6 and 100)
COLOR= <i>color</i>	Color of the font (see below)
ITALIC	Displays the string in italic characters
BOLD	Displays the string in bold characters
HIDE	Hides the string

Examples:

```
TITLE "Monthly analysis" SIZE=10 COLOR=YELLOW BOLD;
TITLE HIDE;
```

Valid Colors:

Black	Maroon	Green	Olive	Navy	Yellow	Purple	Teal
Gray	Silver	Red	Lime	Blue	Fuchsia	Aqua	White

VARDEF

Syntax:

```
VARDEF varlist / [options];
```

Description:

The VARDEF command can be used to assign a value description, or define the display width, number of decimal places, or missing value for one or several variables.

Options:

LABEL "string"	Variable label
WIDTH= <i>integer</i>	Display width
DECIMAL= <i>integer</i>	Number of decimal places to display
MISSING= <i>real</i>	User defined missing value

Examples:

```
VARDEF AGE / LABEL "Age of the child" WIDTH=5 DECIMAL=2;
VARDEF SEX RELIGION INCOME / MISSING = -9;
```

VLABELS

Syntax:

```
VLABELS varlist /
real = "string"
[real = "string"];
```

Description:

The VLABELS command can be used to assign alphanumeric strings to values of one or several numeric variables. When more than one variable is specified, the first variable in the list will contain the value labels, and all the other variables will be linked to this first variable.

Example:

```
VLABEL DEPRESSION1 TO DEPRESSION9 /
0 = "Never"
1 = "Sometimes"
2 = "All the time";
```

WEIGHT**Syntax:**

```
WEIGHT [variable];
```

Description:

The WEIGHT command allows you to designate a weighting variable. When the command is used alone, the weighting is turned off.

Example:

```
WEIGHT GROUPNO;           {use values in GROUPNO as weights}
WEIGHT;                   {turns weighting off}
```

WILCOXON**Syntax:**

```
WILCOXON varlist BY varlist [/options];
```

Description:

The WILCOXON matched-pairs signed-ranks test is a procedure used to test whether two related samples have been drawn from the same population. Like the sign test, it computes the difference between the values of the two variables but takes into account the magnitude as well as the direction of the differences. The Wilcoxon signed-ranks test is the nonparametric version of the t-test for paired samples. The probability test performed can be either one- or two-tailed.

Options:

1TAIL		2TAIL	Direction of the test
PANEL			Display the dialog box

WINDOW**Syntax:**

```
WINDOW windowType [options];
```

Description:

The WINDOW command provides complete control of the display of any SIMSTAT window. The various options allow one to control the size and location of each window and to execute various window related actions (tile, cascade, minimize, etc.). The parameter for TOP, LEFT, WIDTH, and HEIGHT options is an integer value between 0 and 100 expressing a percentage of the screen height and width.

Options:

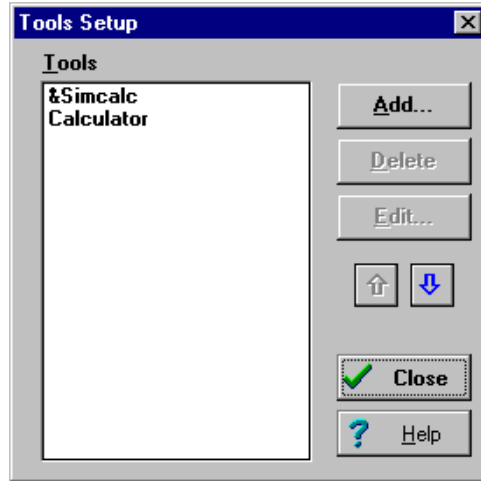
	Window types
DATA	Data spreadsheet window
NOTEBOOK	Notebook/Statistical results window
SCRIPT	Script/log window
CHART	Charts window
ALL	All four windows
TOP= <i>integer</i>	Vertical position of the window's upper left corner (0 to 100)
LEFT= <i>integer</i>	Horizontal position of the window's upper left corner (0 to 100)
WIDTH= <i>integer</i>	Relative width of the window (0 to 100)
HEIGHT= <i>integer</i>	Relative height of the window (0 to 100)
NORMAL	Returns the active window to its size and position before you chose the Maximize or Minimize command.
MINIMIZED	Reduces the active window to an icon.
MAXIMIZED	Enlarges the active window to fill the available space.
CASCADE	Arranges open windows in an overlapped fashion
TILE	Arranges multiple opened windows so they fit next to each other on the desktop and do not overlap.

Examples:

```
WINDOW ALL MINIMIZED;  
WINDOW CHART TOP=10 LEFT=10 WIDTH=50 HEIGHT=50;
```

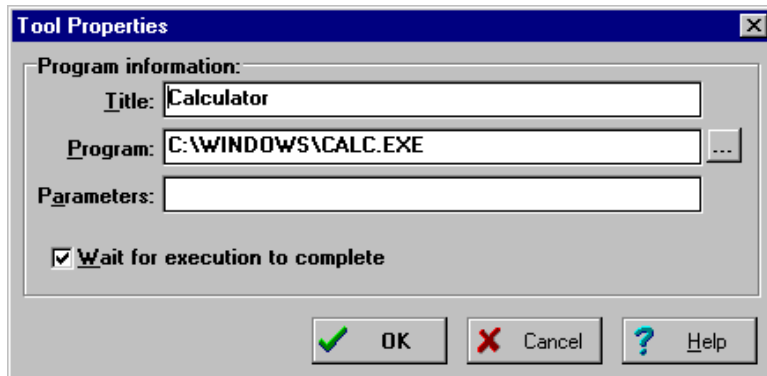

10 - CUSTOMIZING THE TOOLS MENU

The TOOLS pull-down menu can be used to run some SIMSTAT add-in modules or to transfer execution to an external Windows or DOS application such as a file manager, a calculator, or even a word processor. To add programs to, delete programs from, or edit programs on the TOOLS menu, choose the TOOLS SETUP command from the TOOLS menu. The following dialog box will appear.



To add a program to the TOOLS menu

- Choose **Add**. SIMSTAT displays the Tool Properties dialog box, where you specify information about the program to be added.



Title - Enter a name for the program you are adding. This name will appear on the TOOLS menu. You can add an accelerator to the menu command by preceding that letter with an ampersand (&).

Program - Enter the location of the program you are adding. You must include the full path to the program. Click the “...” button to search your drives and directories to locate the path and file name for the program.

Parameters - Enter parameters to pass to the program at startup. For example, you might want to pass a file name when the program launches.

To delete a program from the TOOLS menu

- Select the program to delete.
- Click on the **Delete** button.

To change a program on the TOOLS menu

- Select the program to change.
- Choose the **Edit** button. SIMSTAT displays the Tool Properties dialog box with information for the selected program.
- Modify the program property and click on the **OK** button to save those new properties.

11 - SETTING THE PROGRAM PREFERENCES

The PREFERENCES command gives access to a multi-page dialog box where you can set global options affecting the program's working environment, file handling, and printing. The current section provides a description of each option available in the Preferences dialog box.

ENVIRONMENT PAGE

Show toolbar - This option turns the main window's toolbar at the top of the screen on or off. Turn off the toolbar to gain free work space.

Show status bar - This option turns the status bar at the bottom of the screen on or off. You can gain extra free work space by hiding the status bar.


Display hints - This option allows you to toggle the display of Help Hints.

Beep on error - This option allows you to determine whether or not the computer sounds a beep when an error message appears on the screen.

Show notebook scroll bars - This option allows you to specify whether to display scroll bars on the notebook. Displaying scroll bars allows you to scroll the output horizontally or vertically with the mouse.

Hide menu items unrelated to the active window - This option allows you to choose between two different menu systems: a task specific menu system where only menus related to the active window are shown, and a comprehensive menu system where all menus are displayed even if unrelated to the active window. While the task specific menu system may facilitate the location of a specific command by displaying fewer menu commands at a time, performing a task related to another window requires you to first activate this window in order to gain access to the proper menu. For example, if you are browsing through the notebook and want to filter the data file, you will need to activate the data window before accessing the DATA menu where the FILTER command is located. The full menu system allows you to perform any action related to any of the 4 window types without changing the active window.

Save and restore desktop information - Select this option to keep information about the sizes and locations of the 4 windows at the moment you leave SIMSTAT. The next time you run the program, the locations and sizes of the 4 windows will be restored.

 To make the starting locations and size of those windows permanent:

- Enable this option.
- Adjust the size and location of the 4 windows to your preference.
- Quit SIMSTAT
- Restart SIMSTAT
- Disable this option to prevent overwriting previously saved information.

Display notebook after analysis - When this option is enabled, the Notebook window automatically becomes the active window after an analysis command is performed and its results added to the notebook. (If the Display Graph Window option is enabled (see below) and new charts are created during the analysis, the chart will become the active window).

Display graph window on graph creation - When this option is enabled, the Chart window automatically become the active window after a new chart is created.

Activate script recording on start up - Selecting this option enables the RECORD script feature. This feature automatically generates script commands for actions undertaken using pull-down menus and dialog boxes, and appends those commands to the end of the script window.

DIRECTORIES/BACKUP TAB

Data files directory - By default, when the DATA | OPEN command is activated, the program looks in the drive and directory from which the program was started. This option allows you to specify another default drive and/or directory to be used by the program.

Output files directory - The Output Files directory specifies the name of the directory where both notebook files and chart files are located. If no directory information is provided, the dialogs of the OPEN and SAVE AS commands start on the disk and directory from which SIMSTAT was started.

Script files directory - The Script Files directory specifies the name of the directory where the Script files are located. If no directory information is provided, the dialogs of the OPEN and SAVE AS commands start on the disk and directory from which SIMSTAT was started.

Backup compression factor - This option allows you to set the compression factor to be used by SIMSTAT when creating archive copies of data files. You can set the compression factor to a numeric value between 0 and 9. When set to 0, files are simply stored in the archive. Setting this compression factor to 9 gives the best compression ratio but is also the slowest compression. The default value is set to 6.

Keep a session backup - When this option is enabled, SIMSTAT automatically creates a temporary compressed copy of a data file upon its opening. This feature is especially useful to cancel all data transformation or editing performed during a session and restore the file as it was when you opened it. It is also possible to refresh this temporary backup by using the DATA | SESSION BACKUP | REFRESH option. This will ensure that all modifications to the data file made so far will not be lost if you later decide to revert to a previous version of the data file. If you quit SIMSTAT or open another file, this temporary backup file is deleted.

PRINTING PAGE

The printing page allows you to set various printing options for both text and graphic outputs.

Header - Use this header option to specify a line of text that will appear at the top of each analysis when those analyses are sent to the printer. The header is composed of 3 strings. All text that appears in the **LEFT** box will be printed on the left margin. All text appearing in the **RIGHT** box will be printed flush right, while all text entered in the **CENTER** will be printed in the center of the page.

Special codes can be inserted in any of these boxes to print specific information:

Code	Effect
\$p	This code will be replaced by a page number.
\$d	Put this code in the title to print the date of the page creation.
\$t	Print the time of the page creation.
\$f	This code will be replaced by the original data file name.
\$\$	Print a dollar sign.

Text printing mode - You can choose among 3 different printing methods for printing the content of the notebook.

- **One analysis** - When this option is enabled, every analysis will start printing at the top of a new page.
- **Smart** - When you select this option, the program will try to fit several analyses on a single page. If an analysis cannot be entirely printed in the remaining space, SIMSTAT will start printing at the top of a new page.
- **Economy** - Selecting this option will print every analysis one after the other. Even if an analysis cannot be entirely printed in the remaining space, the first lines will be printed on the current page, while the remaining lines will be printed on a new page.

Left margin - This option allows you to increase the left margin of the printed analysis outputs.

Maximum width - This option lets you specify the maximum width of analysis results. By default, this value is set to 90 characters. You can increase this value up to 254 characters so that larger correlation matrix and contingency tables could be printed.

Font size - This option allows you to increase or decrease the size of the font used to print the content of the notebook. A smaller font size allow you to print more information per page.

Chart printing - The chart printing option allows you to specify how many charts should be printed on a single page. When this option is set to 1 or 4 charts per page, the printing page orientation is automatically set to landscape, while setting the number of charts per page to 2 will force the charts to be printed in portrait mode.

DATA PAGE

The Data page allows you to specify information regarding the management of data files (confirmation of operations, etc.)

The Confirmation section provides check boxes to mark the actions you want SIMSTAT to prompt you to confirm.

Record deletion - Enabling this option causes SIMSTAT to prompt you to confirm the deletion of a record made with the DELETE RECORD command from the DATA menu.

Variable deletion - Enabling this option causes SIMSTAT to prompt you to confirm the deletion of a variable or a variable list specified using the DELETE VARIABLES command.

Variable transformation - Most commands from the Transformation submenu allow you to overwrite values of an existing variable. Enabling this option causes SIMSTAT to prompt you to confirm the overwriting of those values.

Variable creation - Some transformations may require the creation of one or several new variables. Enabling this option causes SIMSTAT to prompt you to confirm the creation of new variables.

The current data file format used by SIMSTAT requires you to determine in advance the physical size of a new variable in the data file. This section allows you to specify the default width and number of decimal places of new numeric variables created during a transformation or an analysis.

Width - The width of a numeric variable should be set to the maximum width that a value can have. The maximum width for the numeric type is 19.

Number of decimals - The number of decimal positions if the field type is numeric. Note that the length of a numeric field that contains decimals includes the decimal point, a leading zero, and an optional sign. The minimum length for a numeric field that contains one decimal position is therefore 3 (unsigned) or 4 (signed). The maximum number of decimals permitted is 16.

The DATA grid section provides two options that allow you to specify how the datasheet will display the content of your data file.

Display font - This option allows you to change the font used to display the data in the data grid. By choosing a different font size, it is possible to increase or decrease the number of rows and columns displayed on a single screen.

Minimum cell display width - While it is possible to individually adjust the display width of each variable in a data file, this option allows you to override these values by setting a minimum display width for the entire data file. This option has been implemented in order to circumvent a performance problem associated with the grid in

the 16 bit version of SIMSTAT. This speed problem occurs when working with large data files involving several hundred variables. Increasing this value reduces the number of columns displayed and thus increases the redrawing speed of the grid.

APPENDIX A - XBASE SYNTAX AND FUNCTIONS

The following section provides a description of xBase syntax rules used in the FILTER, SORTING and COMPUTE commands and a detailed description of each xBase function.

NOTE: This appendix was reproduced from the Apollo v2.0 user's manual with the authorization of Successware 90. Inc.

Expression Operators and Rules

Operators used in xBase expressions are standard in every xBase dialect.

String Operators

- + Joins two strings. Trailing spaces in the strings are placed at the end of each string.
- Joins two strings and removes trailing spaces from the string preceding the operator and places them at the end of the string following the minus sign operator.

Numeric Operators

- + Addition
- Subtraction
- * Multiplication
- / Division
- ^ Exponentiation (or **)

Relational Operators

- = Equal to
- == Exactly equal to
- <> Not equal to
- # Not equal to
- != Not equal to
- < Less than
- > Greater than
- <= Less than or equal to
- >= Greater than or equal to
- \$ Is contained in

Logical Operators (Notice the periods surrounding the operator)

.AND.	both expressions are true
.OR.	either expression is true
.NOT.	either expression is false

Evaluation Order

When more than one type of operator appears in an xBase expression, the order of evaluation is as follows:

Expressions containing more than one operator are evaluated from left to right. Parentheses are used to change the evaluation order. If parentheses are nested, the innermost set is evaluated first.

Numeric operators are evaluated according to generally accepted arithmetic principles:

operators contained in parentheses
exponentiation
multiplication and division
addition and subtraction

Order of evaluation may be altered with parentheses:

$$\begin{aligned}3+4*5+6 &= 29 \\(3+4)*5+6 &= 41 \\(3+4)*(5+6) &= 77\end{aligned}$$

Logical operators are evaluated as **.NOT.** first, **.AND.** second, and **.OR.** last. Logical evaluation order may also be altered with parentheses. In multiple conditional expressions that contain the **.NOT.** operator, always use parentheses to enclose the **.NOT.** operator with the expression to which it applies.

Supported Xbase functions

The following xBase functions are supported in the SORT RECORDS and FILTER RECORDS command

NOTE: Memo field names are not allowed in SIMSTAT xBase expressions.

ALIAS()

Returns the Alias name of the current work area as a string.

ALLTRIM (String)

Trims both leading and trailing spaces from a string. The string may be derived from any valid xBase expression.

```
ALLTRIM("  Provalis  ") returns 'Provalis'.
```

AT (SearchString, TargetString)

Determine whether a search string is contained within a target. If found, the function returns the position of the search string within the target string (relative to 1). If not found, the function returns 0 (zero).

```
AT("gh", "defghij") returns 4.
```

CHR (Val)

Converts a decimal value to its ASCII equivalent.

```
CHR(83) returns 'S'
```

CTOD (String)

Converts a character string into an xBase date. The string must be formatted according to the Windows date format settings.

```
CTOD("12/31/94")
```

DATE ()

Returns the system date (today). Use DTOC(DATE()) to retrieve today's date formatted according to the Windows settings.

DAY (DateField)

Returns the day portion of an xBase date as an integer.

DELETED ()

Returns True if the record is deleted and False if not deleted.

DESCEND (String)

An xBase function that inverts a key value using 2's complement arithmetic. The result of the operation is the arithmetic inverse of the key value. When inverted keys are sorted in ascending sequence, the result is in descending order. A filter expression could be

```
DESCEND(DTOS(billdate)) + CUSTNO
```

DTOC (DateField)

Converts an xBase date into a character string formatted according to the Windows settings. For example, if the date format was American and the date field contained March 21, 1995, DTOC(datefield) would return '03/21/1995'.

DTOS (DateField)

Converts an xBase date into a string formatted according to standard xBase storage conventions (CCYYMMDD). For example, December 21, 1993 would be returned as '19931221'. Indexes that contain date elements should use the DTOS () function, which naturally collates into oldest date first.

EMPTY (Field)

Reports the empty status of any xBase field. Character and date fields are empty if they consist entirely of spaces. Numeric fields are empty if they evaluate to zero. Logical fields are empty if they evaluate to False.

Memo fields that contain no reference to a memo block in the associated memo file are empty.

IF (Logical, True Result, False Result)

This is the immediate if function. If the Logical expression is true, return the True result, otherwise return the False result. The types of the True Result and the False Result must be the same (i.e., both numeric, or both strings, etc.) The logical expression must of course evaluate as True or False.

```
IF (DATE() - CTOD("12/31/93") > 0, "This Year", "Last Year")
```

IIF (Logical, True Result, False Result)

Supported exactly like `IF ()` as noted above.

INDEXKEY ()

Returns the current index key as a string. (Same as `ORDKEY ()`).

LEFT (String, Length)

Returns the leftmost characters of the expression for the defined length.

`LEFT("xyzabc", 3)` returns 'xyz'.

LEN (Expression)

Returns the length of the expression result as an integer.

LOWER (String)

Converts the string expression into lower case.

MONTH (DateField)

Returns the month portion of an xBase date as an integer.

ORDER ()

Returns the current index order as an integer.

ORDKEY ()

Returns the current index key as a string. (Same as `INDEXKEY ()`)

PADC (String, Length, Character)

Centers the passed string between a number of the passed character to make the string the specified length.

`'[' + PADC("Scott", 9, "-") + ']'` returns '['--Scott--']'.

PADL (String, Length, Character)

Pads the passed string to the specified length with the specified characters. If the string is longer than the value specified by Length, the string is truncated to this length.

```
'[' + PADL("Scott", 8, "*" ) + ']' returns '***Scott'.  
'[' + PADL("Loren Scott", 8, " " ) + ']' returns '[Loren Sc]'.
```

PADR (String, Length, Character)

Pads the passed string to the specified length using the specified character. If the string is longer than the value specified by Length, the string is truncated to this length.

```
'[' + PADR("Scott", 8, " " ) + ']' returns '[Scott  ]'.  
'[' + PADR("Loren Scott", 8, " " ) + ']' returns '[Loren Sc]'.
```

RAT (SearchString, TargetString)

Determine whether a search string is contained within a target, starting from the right side of the target string. If found, the function returns the position of the search string within the target string (relative to 1). If not found, the function returns 0 (zero).

```
RAT( "ab", "abzaba" ) returns 4.
```

RECCOUNT ()

Returns the number of records in the table as a long integer.

RECNO ()

Returns the current physical record number as a long integer.

RIGHT (String, Length)

Returns the rightmost characters of the expression for the defined length.

```
RIGHT("xyzabc", 3) returns 'abc'.
```

SELECT ()

Returns the workarea number for the current workarea as a long integer.

SPACE (Length)

Returns a string consisting entirely of spaces for the defined length.

STOD (String)

The inverse of `DTOS()`. `STOD()` converts a string formatted according to standard xBase storage conventions (CCYYMMDD) to an xBase Date formatted according to the Windows settings.

STR (Number, Length, Decimals)

Converts a number into a right-justified string with decimal digits following the decimal point. The total length of the string is defined by the length parameter. `STR(RECNO(), 5, 0)` is a common indexing element that ensures creation of unique keys if appended to another field element.

An index key using this expression could be built with `NAME + STR(RECNO(), 5, 0)`

If the decimals parameter is omitted, the function defaults to zero decimal places. If the length parameter is omitted as well, the length of the result is the length of the field.

STRZERO (Number, Length, Decimals)

Converts a number into a, zero-padded right justified string with decimals digits following the decimal point. The total length of the string is defined by the length parameter.

```
STRZERO( 1234, 10, 2 ) returns '0001234.00'
```

If the decimals parameter is omitted, the function defaults to zero decimals. If the length parameter is omitted as well, the length of the result is the length of the field.

SUBSTR (String, Start, Length)

Returns a portion of the string expression starting at the defined start location for the defined length..

```
SUBSTR('xyzabcd', 3, 4) returns 'zabc'.
```

TIME ()

Returns the system time as a string in the form HH:MM:SS.

TRANSFORM (Expression, Picture)

Transform converts strings and numeric values into formatted character strings. The function transforms the result of the first expression in accordance with the second picture string.

The picture string is made up of two parts. The first part is the Function string and it is optional for both strings and numeric values (as long as the second Template string is present).

A character string transformation picture may consist of only a Function string or only a Template or both.

A numeric picture must contain a Template string; the Function string is optional.

A logical value must contain only a Template string with Template characters L or Y.

The Function string consists of a leading @ character followed by one or more formatting characters. If the Function string is present, the @ character must be the first character in the picture string with its formatting characters immediately following and it may not contain spaces.

If a Template string exists as well, it follows the Function string. A single space separates the Function string and the Template string.

Function string characters allowed for numeric values are:

- B** left justify;
- C** display CR after positive numbers;
- X** display DR after negative numbers;
- Z** blank a zero value;
- (** enclose negative numbers in parentheses.

Function string characters allowed for strings are:

- R** inserts unassigned template characters;
- !** converts all alpha characters to upper case.

The @R Function requires a Template; the ! Function does not.

The Template string describes the format on a character by character basis. The Template string is made up of special characters which have specific results and optional unassigned characters which either replace characters or are inserted in the formatted string depending upon the absence or presence of the @R Function string.

Template assigned characters are as follows:

A,N,X,(,# are place holders and are interchangeable;

L displays logical values as T or F;

Y displays logical values as Y or N;

! converts the corresponding character to upper case;

, (comma) or a space (in Europe) in a numeric template separate the elements of a number;

. (period) or **,** (comma - in Europe) in a numeric template specify the decimal position;

***** fills leading spaces with asterisks in a numeric template;

\$ as the leading character in a numeric template results in a floating dollar sign being placed in front of the formatted number.

Example: Where "phone" is a character field holding a phone number with no formatting characters.

```
'transform(phone, "@R (###) ###-####")' returns '(909) 699-6776'.
```

If the formatting characters were actually present in the field, the "@R" function would be omitted

For numeric fields,

```
'transform(123456.78, "$9,999,999.99")' returns '$123,456.78'.
```

TRIM (String)

Removes trailing spaces from the string expression.

UPPER (String)

Converts the string expression into upper case. Character fields used in index expressions should always be converted to upper case to insure correct collating sequence.

VAL (String)

Converts a string of numeric characters into its equivalent numeric value. The conversion stops at the first non-numeric character encountered (or the end of the string).

```
VAL("123ABC") returns a value of 123.
```

YEAR (DateField)

Returns the year portion of an xBase date as an integer.

APPENDIX B - REFERENCES

STATISTICAL TEXTBOOK

AGRESTI, A. & FINLAY, B. (1986). Statistical methods for the social sciences, 2nd. edition. San Francisco: Dellen-Macmillan.

FERGUSON, G.A. (1989). Statistical analysis in psychology & education, sixth edition. New York: McGraw-Hill.

TABACHNICK, B.G., & FIDELL (1989). Using multivariate statistics, 2nd. edition. New York: Harper Collins.

WINER, B.J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

COMMON PROBLEMS IN STATISTICAL ANALYSIS

ALLISON, D.B., GORMAN, B.S., PRIMAVERA, L.H. (1993). Some of the most common questions asked of statistical consultants: Our favorite responses and recommended readings. Genetic, Social, and General Psychology Monographs, 119(2), 153-185.

COHEN, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

OAKES, M. (1986). Statistical Inference. New York: Wiley.

MEASURES OF ASSOCIATION

GIBBONS, J. D. (1993). Nonparametric measures of association. Beverly Hill: Sage Publication.

HILDEBRAND, D.K., LAING, J.D., & ROSENTHAL, H. (1977). Analysis of ordinal data. Beverly Hill: Sage Publication.

LIEBETRAU, A.M. (1983). Measures of association. Beverly Hill: Sage Publication.

REYNOLDS, H.T. (1977). Analysis of nominal data. Beverly Hill: Sage Publication.

NONPARAMETRIC STATISTICS

CONOVER, W.J (1980). Practical nonparametric statistics 2nd. Ed. New York: John Wiley & Sons.

HOLLANDER, M. & WOLFE, D. (1973). Nonparametric statistical methods, New York: John Wiley & Son.

SIEGEL, S (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

INTER-RATER AGREEMENT MEASURE

BRENNAN, R.L. & PREDIGER, D.J. (1981). Coefficient kappa: Some uses, misuses, and alternatives, *Educational and Psychological Measurement*, 41, 687-699.

COHEN, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20, 30-46.

KRIPPENDORFF, K. (1970). Bivariate agreement coefficients for reliability of data. In E.F. Borgatta and G.W. Bohrnstedt (Eds.). *Sociological methodology: 1970*. San Francisco: Jossey-Bass.

SCOTT, W.A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19, 321-325.

MULTIPLE COMPARISON PROCEDURES

JACCARD, J., BECKER, M.A., & WOOD, G. (1984). Pairwise multiple comparison procedures: A review. *Psychological Bulletin*. 96, 589-596.

TOOTHAKER, L.E. (1993). *Multiple comparison procedures*. Beverly Hill: Sage Publication.

MULTIPLE REGRESSION

DARLINGTON, R.B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161-182.

COHEN, J. & COHEN, P. (1983). *Applied Multiple Regression/Correlation analysis for the Behavioral Sciences*. 2nd Ed. Hillsdale, N.J.: Lawrence Earlbaum.

DRAPER, N.R. & SMITH, H. (1981). *Applied Regression Analysis*. 2nd Edition. New York: John Wiley & Sons.

PEDHAZUR, E.J (1982). *Multiple Regression in Behavioral Research*, 2nd Edition. New York: Holt, Rinehart and Winston.

GENERAL LINEAR MODEL & ANALYSIS OF VARIANCE/COVARIANCE

- COHEN, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70, 426-443.
- COHEN, J. & COHEN, P. (1983). *Applied Multiple Regression/Correlation analysis for the Behavioral Sciences*. 2nd Ed. Hillsdale, N.J.: Lawrence Earlbaum.
- IVERSEN, G.R., & NORPOTH, H. (1976). *Analysis of Variance*. Beverly Hill: Sage Publication.
- PEDHAZUR, E.J (1982). *Multiple Regression in Behavioral Research*, 2nd Edition. New York: Holt, Rinehart and Winston.
- OVERALL, J.E., & SPIEGEL, D.K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, 72, 311-322.

ITEM ANALYSIS

- CROCKER, L., & ALGINA, J. (1986). *Introduction to classical & modern test theory*. Fort Worth: Harcourt Brace Jovanovich College Publishers.
- EBEL, R.L. (1965). *Measuring educational achievement*. Englewood Cliffs: Prentice-Hall.
- HENRYSEN, S. (1971). Gathering, analyzing and using data on test items. In R.L. Thorndike (Ed.). *Educational measurement* (2nd Edition). Washington, D.C.: American Council on Education.
- STREINER, D.L., & NORMAN, G.R. (1989). *Health measurement scales: A practical guide to their development and use*. Oxford University Press.

SINGLE-CASE EXPERIMENTAL DESIGN

- HERSEN, M., & BARLOW, D. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press.
- GRESHAM, F.M. (1991). Moving beyond statistical significance in reporting consultation outcome research. *Journal of educational and psychological consultation*, 2(1), 1-13.
- MICHAEL, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, 7, 647-653.
- SIDMAN, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Author's Cooperative Press.

RELIABILITY ANALYSIS

- CARMINES, E.G., & ZELLER, R.A. (1979). Reliability and validity assessment. Beverly Hill: Sage Publication.
- NUNNALLY, J.C. (1964). Educational measurement and evaluation. New-York: McGraw-Hill.

BOOTSTRAP SIMULATION

- DIACONIS, P., & EFRON, S. (1983, May). Computer intensive methods in statistics. *Scientific American*, 116-130.
- EFRON, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap, and other resampling methods. *Biometrika*, 68, 589-599.
- EFRON, B., & GONG, G. (1983). A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician*, 37, 36-48.
- EFRON, B., & TIBSHIRANI, R.J. (1993). An introduction to the bootstrap. New York: Chapman & Hall.
- MOONEY, C.Z., & DUVAL, R.D. (1993). Bootstrapping: A nonparametric approach to statistical inference. Beverly Hill: Sage Publication.
- PÉLADÉAU, N., & LACOUTURE, Y. (1993). SIMSTAT: Bootstrap computer simulation and statistical program for IBM personal computers. *Behavior Research Methods, Instruments, & Computers*, 25(3), 410-413.
- STINE, R. (1989). An introduction to bootstrap methods: Examples and ideas. *Sociological Methods and Research*, 8(2&3), 243-290.
- WASSERMAN, S., & BOCKENHOLT, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, 26(2), 208-221.

APPENDIX C - LIMITATIONS

Maximum number of variables	1022
Maximum number of cases	320,000 or more
Maximum number of bootstrap resampling	32,000
Maximum number of notebook pages	16,320
Maximum number of charts	16,320
Maximum number of sections in a notebook:	100
Maximum number of nominal values (Frequency, Crosstab Breakdown, Oneway, Kruskal-Wallis)	255
Maximum number of factors/covariates (GLM ANOVA/ANCOVA)	5
Maximum number of predictors in multiple regression	40
Maximum number of variables in reliability analysis	90
Maximum number of variables in item analysis	250

APPENDIX D - TECHNICAL SUPPORT

When contacting Provalis Research for technical support by fax or e-mail, be sure to include your product name and version. Include a description of the problem, including all the steps needed to replicate it and, if applicable, a complete program source code and data (in the smallest form possible). You can contact Provalis Research:

Phone: 514-899-1672 (collect call not accepted)
FAX: 514-899-1750
Compuserve E-mail: simstat
Internet E-Mail: support@simstat.com
Internet Web site: <http://www.simstat.com>
Standard mail: Provalis Research
2414 Bennett Street
Montréal, QC
H1V 3S4
CANADA

INDEX

A

ACF plot 146, 228
Agreement measures 97-99
Alpha
 Bootstrap analysis 69-70, 72, 183, 217
 Cronbach's Alpha 100-107, 137-139, 221, 247
ANCOVA 90-93, 196
 References 250-251
ANOVA
 Friedman test 89, 195
 GLM ANOVA/ANCOVA 90, 93, 196
 Kruskall-Wallis 110, 202
 Multiple regression 119, 120, 208
 ONEWAY 130, 210
 Regression analysis 133, 220
 References 250-251
Append data 42, 181
Archival backup of data files 45
Archives 22, 45
ASCII files 39, 42, 191
Autocorrelation 146-148
AVI files 180, 213
Axis 158, 159, 162-164, 186, 193

B

Background color 162, 224
Backup 3, 45, 236
Backward selection 119, 208
Barchart
 Crosstabulation 78-79, 191
 Frequency 86-87, 195
 Inter-raters analysis 98-99, 200
Binomial test 67, 181
Bitmap 10, 165-166, 214
BMP files 10, 165-166, 214
Bootstrap analysis 68-73, 182, 183, 196, 217
 References 252
Box-&-Whisker plot 74, 87, 185, 195
Breakdown analysis 74, 185

C

Caseplot
 GLM ANOVA/ANCOVA 92, 197
 Multiple regression 120, 220
 Regression 133, 220
Charts 153-166
 3D View 162
 Axis scaling and grid 159
 Creating specific charts 154

 Customizing charts 158-164
 Exporting charts 165
 Data point values 160
 Legend 160
 Navigating in the chart window 155
 Setting global options 160
 Titles and axis labels 158
 Zooming in and out 163
Chi-square 79, 129, 186, 190
CHOOSE X-Y 14, 65
CHX files 19, 157, 211
Clipboard 6, 10, 155, 165, 170
Cluster analysis 66, 187
Cohen's Kappa 83, 97, 187, 200, 217
Compression 45, 236
Conditional transformation 55, 188
Confidence intervals
 Bootstrap analysis 68-70, 72, 182, 183, 217
 Correlations 75, 190
 Frequency analysis 86, 195
 Logistic regression 112-113, 204
 Oneway ANOVA 130, 210
 T-test 149, 227
Confirmation 238
Contingency table 77, 115, 125, 190
Control bars 144-145, 147, 223, 228
Correlation
 Bootstrap 70, 183, 196, 217
 Correlation matrix 75, 190
 Crosstabulation 77, 125, 190, 207
 GLM ANOVA/ANCOVA 92-197
 Item analysis 104, 201
 Logistic regression 112, 204
 Multiple regression 120-220
 Regression analysis 133, 220
 Reliability analysis 136-139, 221
Correspondence analysis 189
Covariance matrix
 Correlation analysis 75, 190
 Factor analysis 81, 193
 GLM ANOVA 90, 196
 Reliability analysis 136-138, 221
Creating a new data file 24
Creating specific charts 153
Crosstabulation 77, 125, 190, 207
 References 249
Cubic regression 136, 220
Cumulative distribution chart 86, 195
Customizing charts 158-164

D

Data distribution	49, 72, 87
Data files	
Backup of data files	45
Creating a new data file	24
Defining variables	27
Entering and Editing Data	31
Exporting data to other applications	40
Filtering records	33
Importing data from other applications	38
Limiting access to data files	43
Merging data files	42
Opening an existing data file	21
Sorting records	36
Data labels (in a chart)	160
dBase files (DBF)	22, 41, 191, 211
Decimal places	
Variables	25, 27-28, 50, 226, 238
Output	64
Charts	159, 186
Defining variables	27, 230
Delete record	31, 238
Descriptive analysis	72, 80, 86, 130, 193
Deviation chart	131, 211
Directories	236
Discrimination index	100, 201
Displaying data point values	160
Diversity indices	193
Dual histogram	150, 227
Dummy recoding	58
Durbin-Watson	
GLM ANOVA/ANCOVA	92, 197
Multiple regression	120, 208
Regression analysis	134, 221

E

Editing axis scaling and grid	159
Editing titles and axis labels	158
Effect coding	59, 90
Eigenvalue	81-84, 189, 193, 212
Encrypted script files	168
Endorsement rate	104
Entering and editing data	31
Error bar chart	130, 149, 210, 227
Error rates	68, 141, 225
Excel files	38-40, 199
Exponential regression	52, 133, 176, 220
Exportation	
Charts	155, 165-166
Data to other applications	40-41
Expression operators and functions	176, 240

F

Factor analysis	81, 193
Filtering records	33, 194
Font	158, 161, 209, 229, 238
Forward selection	119, 208
Frequency analysis	15, 86, 97, 125, 186, 195
Friedman test	89, 195
Full analysis bootstrap	73, 196
Further reading	249-252

G

Gamma	69, 78, 127, 183, 191, 210, 218
Getting help	11
GLM ANOVA/ANCOVA command	90, 196, 250
Global options	160, 225, 235
GO TO command	60

H

Header	225, 237
Help	11
Hints	9, 225, 235
Hierarchical regression	90, 118-119, 208
Histogram	49, 72, 87, 150, 183, 184, 195, 217, 227

I

Image covariance	81, 193
Import data	38, 199
Independent t-test	149, 227
Installation	3
Interaction	
GLM ANOVA/ANCOVA	90, 91, 196-197
Logistic regression	112, 204
Internal consistency	100, 136-137, 221, 251
Inter-item correlation	137, 221
Inter-raters agreement	97, 200, 250
Inverse regression	133, 220
Item analysis	104, 201, 251
Item characteristic curves	104, 201

K

Kendall's tau	69, 78, 127, 183, 210, 218
Kolmogorov-Smirnov test	108, 109, 202, 203
Krippendorff agreement measures	97, 200, 250
Kruskal-Wallis command	110, 202
Kurtosis	49, 69, 74, 86, 100, 125, 182, 193

L

Labels	
Axis labels	160, 162
Value labels	22, 27-30, 38, 41, 45, 230
Legend	160-162
Likelihood ratio	78, 112, 191, 204
Limitations	38, 40, 253
Linear regression	19, 220
Listing cases	111, 204
Listwise deletion	75, 104, 127, 190, 201, 210
Logistic regression	112, 204
Lotus files	38, 40, 199

M

Mann-Whitney test	114, 152, 205
Manual conventions	2
McNemar test	115, 205
Median	49, 69, 74, 86, 182, 183, 193
Median test	116, 206
Running median	224, 228, 144
Split-median trend	224, 144
Median test	116, 206
Memory and data file variables	174
Merging data files	42
Metafiles	6, 165
Missing values	22, 28, 53, 57, 75, 127, 190, 201, 210, 219
Modifying specific chart options	160
Moses test	117, 207
Movie file	180, 213
Moving average	145, 147, 224, 228
Multiple comparison procedures	130, 211, 250
Multiple regression	90, 118-124, 196, 197, 204, 208
References	250
Multiple responses analysis	125, 207

N

Navigating	
in the data windows	23
in the notebook	61
in the chart window	155
in the script window	170
Newman-Keuls	130, 211
Nonlinear regression	133, 220
Nonparametric tests	
Binomial test	67, 181
Friedman test	89, 195
Kolmogorov-Smirnov test	108, 109, 202, 203
Kruskal-Wallis command	110, 202
Mann-Whitney test	114, 152, 205
McNemar test	115, 205
Median test	116, 206

Moses test	117, 207
NPAR matrix	127, 210
One sample chi-square test	129, 186
Runs test	140, 222
Sign test	143, 226
Wilcoxon test	152, 231
Normal probability plot	49, 87, 120, 134, 195, 197, 208, 220
Notebook	4, 60-64, 225
NPAR matrix	127, 210, 185
Numeric recoding	59

O

One sample chi-square test	129, 186
Oneway ANOVA	110, 130, 210, 251
Opening a script file	169
Output management using Tabs	63

P

PACF plot	146, 228
Pack file	31
Pairwise deletion	75, 127, 190, 201, 210
Paradox files	38, 40, 199
Pareto chart	86, 87, 195
Partial autocorrelation	146, 228
Password protection	43
Pearson's r	75, 78, 191
Percentile table	69, 86, 182, 184, 195, 217
Phi	78, 190, 191
Point marker	162
Power estimate	69, 72, 182
Principal components analysis	212
Principal coordinates analysis	212
Printing	19, 215, 237
Normal probability plot	49, 87, 120, 134, 195, 197
Program preferences	225, 235

Q

Quadratic regression	136, 220
Quattro Pro files	38-40, 199
Quitting	207, 217, 235, 236

R

Rank	
Transforming data into ranks	58, 219
Statistics	89, 110, 114, 152, 195, 202, 219
Recoding values	57-59, 219
Record	
Delete record	31, 238
Filtering records	33, 194
Sorting records	36, 227

Record script	64, 171, 236
References	249-252
Regression analysis	
Bootstrap	70, 183, 196, 217
GLM ANOVA/ANCOVA	92-197
Logistic regression	112, 204
Multiple regression	120-220
Regression analysis	133, 220
Reliability analysis	100-103, 136-139, 221, 251
Residual analysis	87, 120, 134, 195, 197, 208, 220
Rotated factor solution	81, 194
Rounding numerical values	64
Running a script	171, 185
Running median	145, 146, 224, 228
Runs test	140, 222

S

Saving script files	172
Scatterplot	133, 221
Residual scatterplot	92, 120, 134, 197, 208, 220
Scheffé	130, 211
Scott's pi	183, 201, 218, 244, 245, 250
SCR files	168, 172, 185
Scree plot	81, 83, 194
Script	167-180
Encrypted script files	168
Expression Operators and Functions	176
Introduction	168
Memory and data file variables	174
Navigating in the script window	170
Opening an existing script	169
Recording a script	171
Running a script	171
Saving a script	172
Syntax Convention	173
SCZ files	168, 169, 172
Seasonality	146
Security	28, 43-44
Seed value	71-73, 182, 183, 196, 218
Sensitivity analysis	141-142, 225
Sets of variables	22, 46, 118
Setting global options	160
Sign test	143, 226
Single-case experimental design	223, 251
Skewness	49, 69, 74, 86, 100, 125, 182, 193
SNB files	19, 211
Sorting	36, 237, 77, 86, 97, 191, 195, 200, 227
Sound files	180, 213
Split-half reliability	139, 221
Spreadsheet	39-41, 199
SPSS	38, 40, 199
Standard regression	118, 119, 208

Startup options	4
Status bar	11, 225, 226, 235
Stepwise regression	118, 119, 121, 208
Symphony files	38-40, 199
Syntax Convention	173
System requirements	3

T

T-test	149, 227
Tab delimited files	38-41, 165, 199
Tabs	39, 60, 63
Tau	69, 78, 127, 183, 191, 210, 218
Technical support	2, 3, 253
Time-series analysis	147, 228
Toolbar	9-11, 225, 235
Tools menu	233
Transformation	50, 57-59, 188
Rank transformation	58, 219
Tukey	130, 211
Tutorial	13-19

V

Value labels	22, 27-30, 38, 41, 45, 230
Variable assignment	18, 65
Variable name	24, 27, 28
Variable sets	45-46
Variable statistics	49
Varimax rotation	81- 85, 194

W

WAV files	180, 213
Wilcoxon test	152, 231
Windows bitmap	165, 214
Windows metafile	165, 214
WMF files	165, 214

X

xBase syntax and functions	240-248
----------------------------	---------

Z

Zooming in and out	163
--------------------	-----

