APPENDIX S

INTRODUCTION TO BASICS OF SPSS

SPSS is a collection of programs specialized for the creation of data sets and for statistical analysis of data. A data file or data set involves observations for one or more variables on a number of cases. For example, to study the relationship between test anxiety and grade point average, researchers could obtain scores (the observations) for test anxiety and grade point average (the variables) on a number of people (each person would be a case to SPSS). One could imagine the results from such a study as a table, with the cases as rows, the variables as columns, and the values in the individual cells as observations. Users familiar with the terminology of database languages may know data sets as databases, cases as records, and variables as fields. The corresponding SQL terms are tables, rows, and columns.

Because SPSS is a mature program that has been around for some time, there are different versions for different computers, and even for the same computer. Although the various versions demonstrate some commonalities, there may be variations in options of commands or other differences that users must attend to. The present material focusses on recent versions of SPSS for Windows, although a few comments will be made about using SPSS on other systems, specifically Unix-based computers.

SPSS BASICS

We begin with some basic features of SPSS, such as how to start SPSS, understanding its displays, creating data files, and so on. We later examine other generally useful features.

Running SPSS

SPSS for Windows can be started in several ways. If the Desktop includes an icon for SPSS, then users can double-click to initiate SPSS. Or it may be necessary to access SPSS from the Start menu, generally some sequence like: *Start* | *Programs* | *SPSS for Windows*. If someone has defined a short-cut key to invoke SPSS (e.g., Ctrl-Alt-s), then typing that short-cut key will initiate SPSS. Once initiated, users can enter SPSS commands and perform a variety of operations relevant to the analysis of data.

On Unix systems, access to SPSS is somewhat less direct. The user first creates a file

containing SPSS commands (e.g., job1.sps), and then invokes SPSS from the command line, with a command of the following sort: *spss -m job1.sps > job1.lst* . This command tells the Unix operating system to invoke SPSS, read and process the commands in the file job1.sps, and then send the output to the file job1.lst.

SPSS Windows

SPSS for Windows involves different windows (or screens) that users can switch between. These are essentially distinct working areas within SPSS and usually serve a defined purpose. The Windows that we will discuss are: the Data Editor window, the Syntax window, and two different output or Viewer windows. To transfer between windows, users can select a specific window from the pull-down menu (see Figure S.1) or can click on the window's icon on Window's taskbar. Windows also provides several shortcut keys for moving between successive windows.

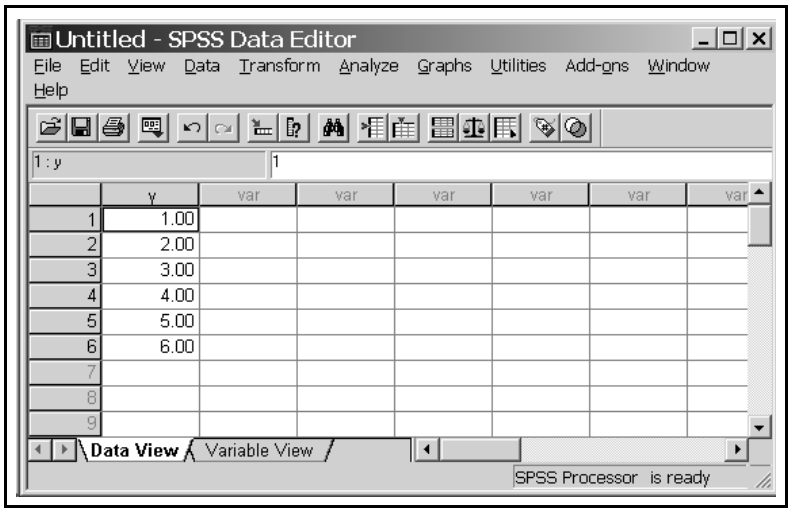

Figure S.1 - SPSS Data Editor Window (Screen).

Data Editor Window

The Data Editor window allows users to enter data, view data that has been created or computed by the user, define variables, and the like. Figure S.1 shows the Data Editor window in SPSS. In this particular example, the user has already typed in values for the variable represented by the first column. The values of 1 ... 6 in the column headed y represent scores and were simply typed into their respective cells. The numbers 1... 6 to the left indicate the rows or cases of the data file.

The default view in the Data Editor window is the Data View (note the highlighted Data View tab at the bottom of Figure S.1). Note also that the empty columns have as headings the label *var*, which is the default variable name when data is entered directly into the data editor window. To change the name and other properties of a variable, select the Variable View tab at the bottom. The resulting screen is shown in Figure S.2 and allows the user to change the name of the variable (double-click on the current name and type the new name), and such other features as the total number of columns to set aside for the variable and the number of decimal places to use. Values of 4 and 2 are generally appropriate for our uses.

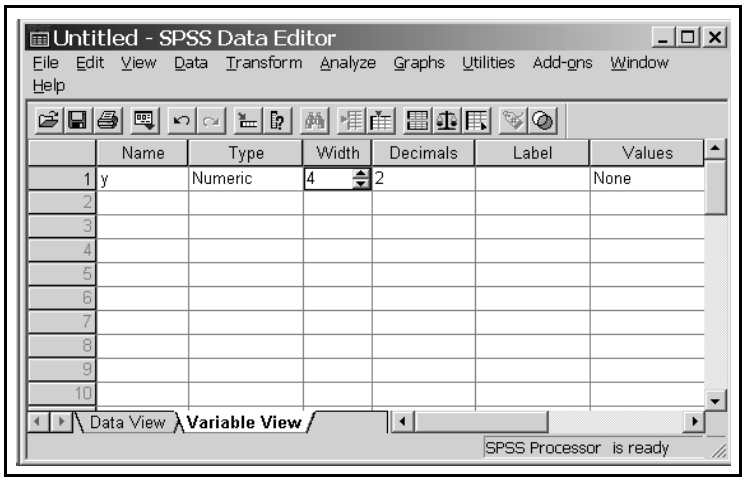

Figure S.2 - Variable View Screen of Data Editor Window.

Syntax Window

The Syntax window is where SPSS commands can be entered and run. This window can be created by choosing *File | New | Syntax*, from the pull-down menu, as shown in Figure S.3. Rather than always showing the actual screen for menu commands, the format of the preceding sentence will often be used; that is, a series of menu choices will be shown italicized and separated by vertical lines (|). Menu choices can be made using the mouse (or other manual device) or by pressing Alt-letter, where Alt is the Alt key and letter is the underlined letter on the menu. Alt-f, for example, opens the File option on the menu.

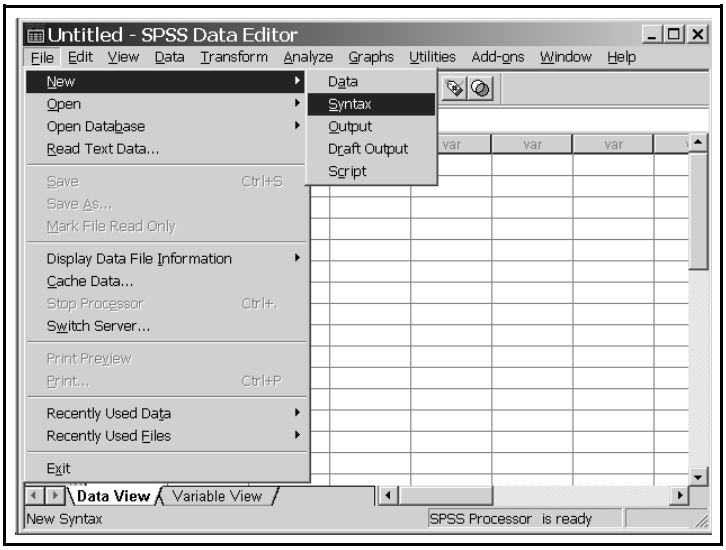

Figure S.3 - Starting a Syntax Window.

The actual syntax window is shown in Figure S.4, along with some SPSS commands that have been typed in by the user. Essentially, SPSS has a vocabulary of terms that allows users to control its action. Sentences are constructed using appropriate terms, and ended with periods. Figure S.4 shows three SPSS commands (commands can extend over more than one line). The first line is the *DATA LIST* command, which can be used to enter data into SPSS (instead of typing it into the Data Editor window). Like most SPSS commands, *DATA LIST* has a number of options. The *FREE* option tells SPSS that the to-be-read data is in free format (i.e., individual scores are separated by one or more spaces), the most unstructured of the data formats used by SPSS. The / is a separator used in SPSS to separate distinct sections of commands; in the case of *DATA LIST*, the separator is followed in our example by a single variable name chosen by the user. To keep SPSS terms and user terms distinct, I will type the SPSS terms in uppercase characters and the user-defined terms in lowercase characters (y in our example). However, SPSS does not require this separation. The period following y denotes the end of this command.

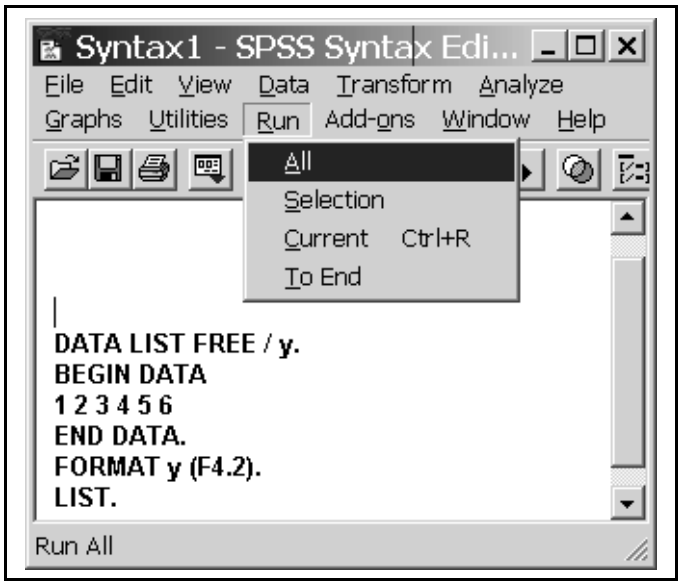

Figure S.4 - SPSS Syntax Window.

The next command is the *BEGIN DATA ... END DATA* command, which spans three lines (it could be more). Anything between *BEGIN DATA* and *END DATA* is treated as data. In this example, SPSS will read the 6 scores as values of y for each of 6 cases (subjects). The *FORMAT* line changes the format for y from the default values (often 8 columns and 2 decimals, denoted as 8.2) to 4 columns and 2 decimals. After this command is invoked (see next section), SPSS would contain a data file identical to that shown in Figure S.1 and the *LIST* command could be used to list the data in the SPSS output screen.

Running SPSS Syntax Commands

Figure S.4 also shows the *RUN* options available via the Syntax window's pull-down

menu. Users can run *All* of the entered commands from the first to last, a previously highlighted *Selection* of commands (selection could be done using the mouse or Shift-Arrowkey combinations), the command under the *Current* position of the cursor (the cursor can be anywhere on the command and need not be at the start), or the commands from the current cursor position *To* the *end* (or last) command in the syntax file. Note that the *Current* option allows use of a short-cut key (Ctrl-r); that is, positioning the cursor anywhere in an SPSS syntax command and pressing Ctrl-r will invoke that command. Ctrl-r can also be used in conjunction with the selection feature; that is, selecting a range of commands and pressing Ctrl-r will invoke the full range of selected commands, and not just the command on which the cursor is positioned.

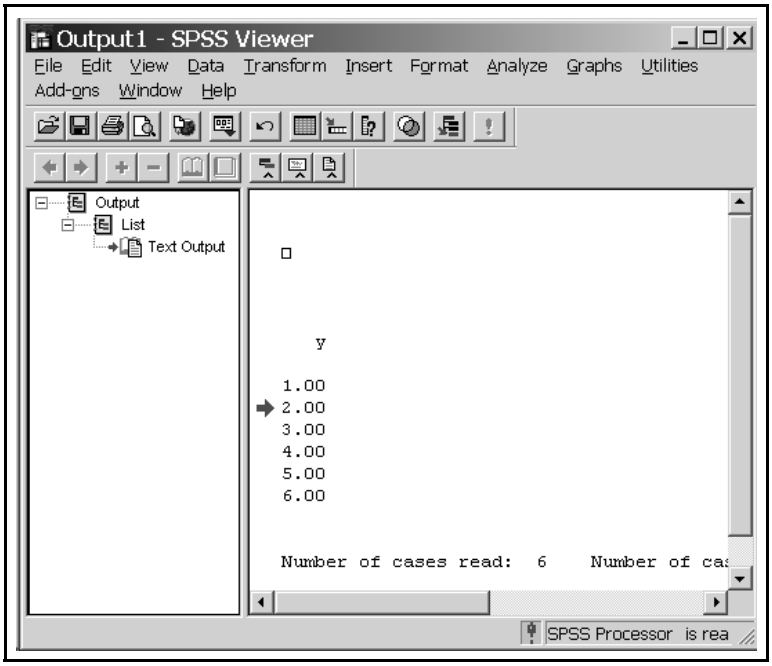

Figure S.5 - SPSS Output Viewer Window.

SPSS Output Window(s)

Figure S.5 shows the default output screen for SPSS; this is a very sophisticated output screen that can be used in powerful ways to control the output display. The left section is a navigation panel that allows users to choose what part of the output they wish to view, and the right panel is the actual output from SPSS. In this example, we see a listing of our 6 scores, just as we requested in the syntax file in Figure S.4.

SPSS also has a cruder output window, which it calls the Draft Output window. This window can be activated by selecting *File | New | Draft Output* on the menus. Because it is simpler, draft output can be easier to modify if users want to copy and paste output into other documents, and then edit the output.

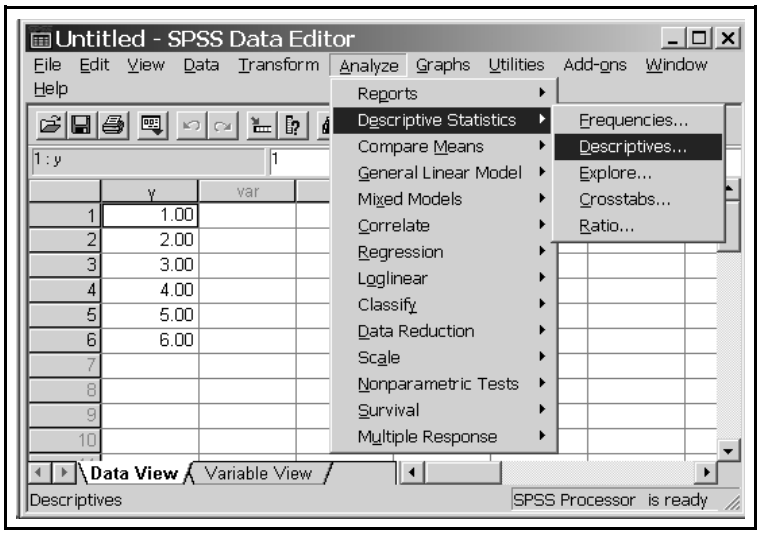

Figure S.6 - Menu Commands for Descriptive Statistics.

Statistical Analysis of Data Sets

Once we have defined our variables and created datasets, then we can conduct statistical analyses, either using the pull-down menus or by typing and running additional commands into the Syntax window. Consider basic descriptive statistics, such as the Mean, Standard Deviation, and Variance of our variable(s). Figure S.6 shows the *Analyze | Descriptive Statistics | Descriptives* sequence needed to invoke the descriptive statistics module.

Figure S.7 shows the actual Descriptives window. When first opened, this window would contain a list of all variables in the current dataset in the left panel; in our simple example, this would simply be the variable y. Users would select the variables for which they want descriptive statistics, and then click on the center arrow. This arrow would have been pointed from the left to the right panel, and clicking on it moves the name of the variable(s) into the right panel. The right panel lists the variables to be analyzed. If satisfied with the default statistics, users could then click on

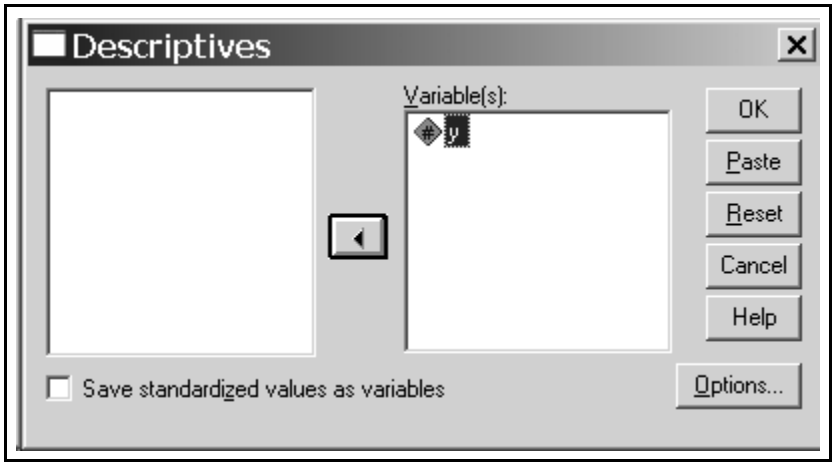

Figure S.7 - Main Window for Descriptives Procedure.

Ok and descriptive statistics for the specified variables would appear in the output screen.

Alternatively, users can select the Options button for Descriptives, and then choose what statistics they wanted and did not want from the dialogue box shown in Figure S.8. Clicking on the boxes will select or de-select different options desired by the user. We want the Mean, Standard Deviation, and Variance; hence, we have unselected Minimum and Maximum (part of the default output) and selected Variance (which is not part of the default). Clicking on *Continue* will return us to the screen in Figure S.7, and clicking on Ok will produce the desired output (barring any errors in our specification of the job).

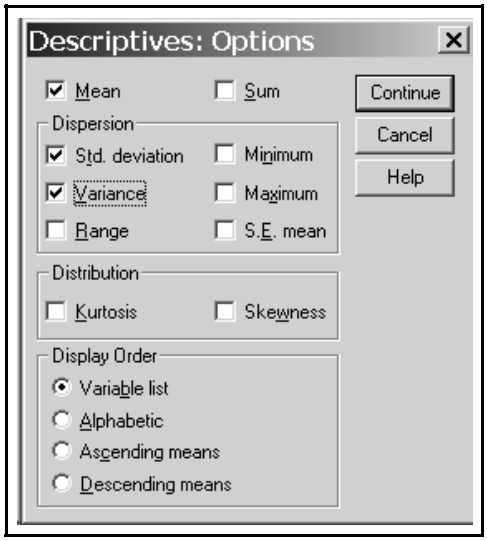

Figure S.8 - Choosing Descriptive Statistics.

Figure S.9 shows the resulting output from the Descriptives command. SPSS reports the Mean, Standard Deviation, and Variance for y, the variable that we had specified in the Descriptives dialogue box.

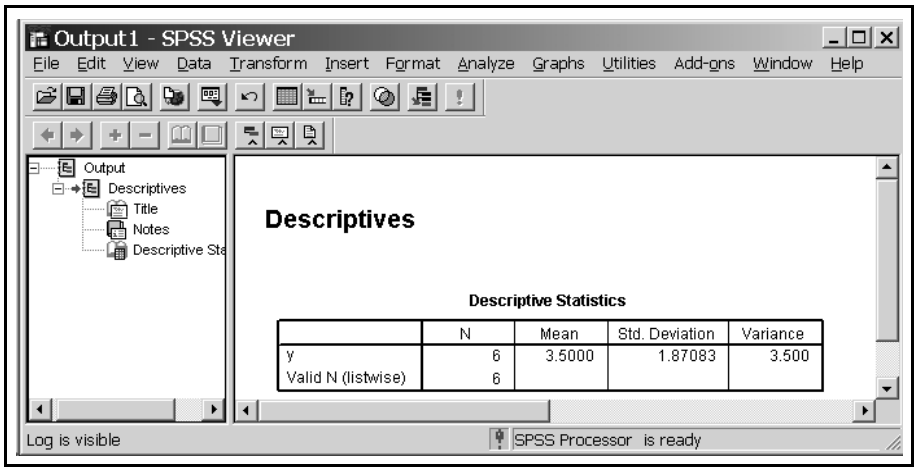

Figure S.9 - Descriptives Output.

This same output would be produced if we typed *DESCRIPTIVES y /STATISTICS = MEAN STDDEV VARIANCE* in the Syntax window and then ran the command. Note that many SPSS commands can be abbreviated to just enough letters to uniquely identify the SPSS keyword. The preceding could have been written simply as *DESCR y /MEAN STDDEV VARI*.

Menu Commands and Syntax Commands

Novice users often wonder why bother with syntax, whereas more sophisticated users generally appreciate the many benefits of syntax. In particular, syntax makes it possible to save

commands in a syntax file, and then retrieve it at a later time to repeat the analysis, perhaps with some slight modifications to the syntax. Moreover, some features of SPSS can only be accessed by syntax (e.g., INPUT PROGRAM, MANOVA).

There is a more intimate relationship between the Menu and Syntax approaches than is obvious to beginning users. The Menu system actually generates syntax commands that are then performed by SPSS. To see this connection, it is necessary to turn on a feature of SPSS that is sometimes not included in the default configuration. Specifically, we need to request SPSS to include a log (i.e., listing) of the SPSS commands in the output. This is handy, in any case, to make clear what commands were used to produce the output. Selecting *Edit | Options* on the main SPSS menu invokes the Options dialogue box in Figure S.10. On the left, lower side, we see the *Display commands in the log* box, which needs to be checked if we want SPSS commands shown in the output. Depending how options are set for your version of SPSS, it may also be necessary to configure the log to be shown (see Figure A1 - 10).

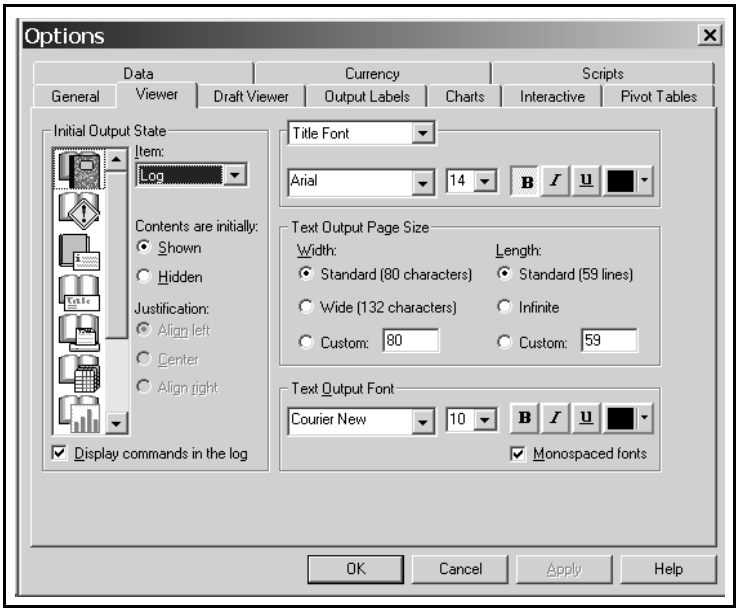

Figure S.10 - Turning on the Display Log Feature.

Figure S.11 shows the output from the Descriptives analysis (via menus) after this change has been made. In addition to the statistical results, SPSS now also shows the

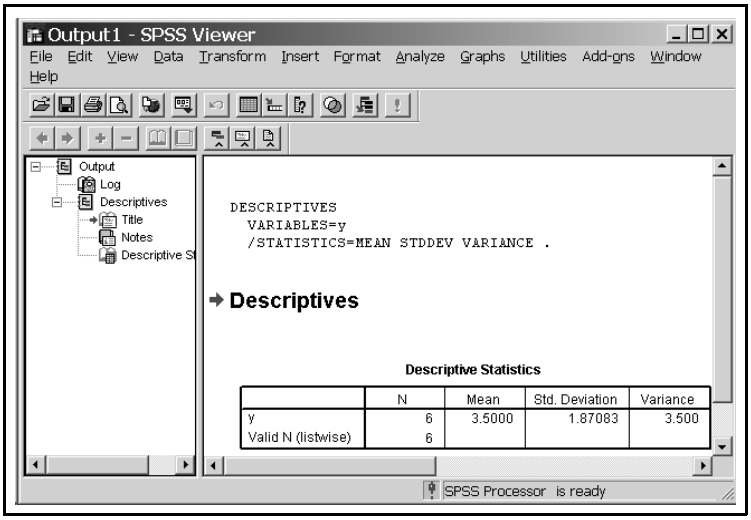

Figure S.11 - Descriptives Output with Syntax Logging Turned On.

SPSS commands that produced this output. The syntax generated using menus is often more complex than syntax that we type directly. This occurs because SPSS does not abbreviate commands and options, and because SPSS includes the syntax for default values that need not be included when commands are typed.

Turning on the syntax log allows users to see the syntax generated by their menu selections (additional changes to Options may be necessary depending on the default configuration of SPSS on the system you are using). This syntax can also be copied and pasted into a syntax file, which could then be saved and retrieved on a future occasion if there was some need to run the commands again, perhaps modified in some way or using a different dataset.

Saving and Retrieving SPSS Files

SPSS allows users to save and retrieve (i.e., open) the contents of various windows. Of particular interest to us is the syntax window, which is how we will generally enter or generate data, and how we will enter (or copy) SPSS commands to perform various statistical analyses. We can also save Data sets in an SPSS format that retains generated or modified variables (see below), variable names, and the like. And it is possible to save Output screens to be examined or printed at a later point in time. The commands to Save and Open files are on the File option of the main menu. The format of the saved file will vary somewhat with the specific window from which it is accessed. From the Syntax window, for example, SPSS would save the file as text and the filename would have a .sps extension at the end. Data sets saved in SPSS format would have a .sav extension.

Users would need to specify the location for the file to be saved or opened from, including the drive and folder name along with the filename. Given the more complex networks that are common in today's computing environment, users might have a number of options (e.g.,, networked drives, floppy or A: drive, hard-drive, removable hard-drive).

FURTHER OBSERVATIONS ON SPSS

Although the preceding is enough to get started in SPSS, a few additional observations might preclude some problems (and open new doors).

Modifying the Draft Viewer to Copy Output

Users who want to copy and edit SPSS output extensively can either learn the commands to manipulate the default graphical viewer, or else use the cruder Draft Viewer, with some modifications. One modification that helps in editing output is to eliminate the characters that SPSS uses to put boxes around the output. Choose *EDIT* | *OPTIONS* from the main menu, and then select the tab for the Draft Output to reach the dialogue box shown in Figure S.12. Note the Cell Separators boxes in the middle, right section of the figure. The Row and Column separators originally contained characters (e.g., | for column separator). Placing a single blank space into these boxes will result in cleaner output that is easier to copy and edit in other programs.

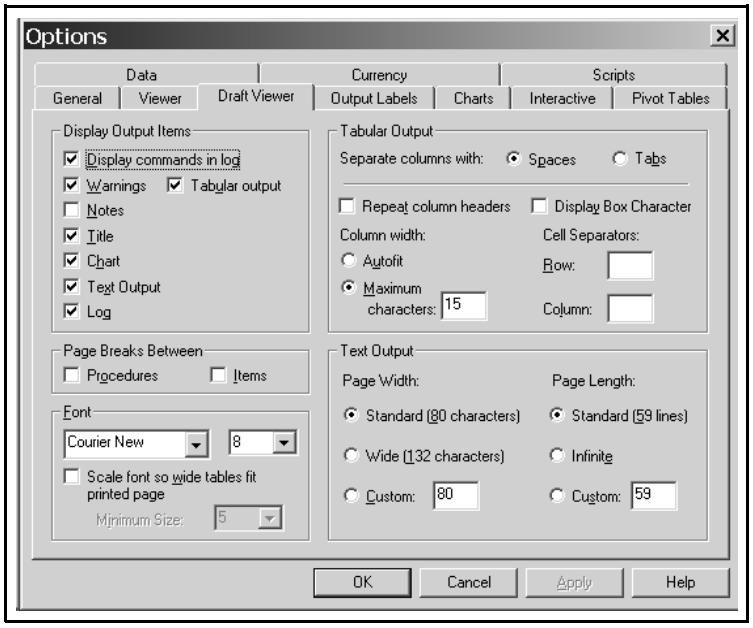

Figure S.12 - Draft Viewer Options.

Variable Names and Variable Lists

Variable names play an important role in SPSS, and there are a number of conventions covering their use. Variable names can be a maximum of 8 characters long. Variables usually begin with a letter (A-Z). After the first letter, variable names normally contain alphabetic (A-Z) or numeric (0-9) characters. Thus, ANXIETY, ANX1, DRUG, and V01 are all valid variable names. DRUGOVERDOSE would not be a legitimate name because it exceeds the 8-character restriction.

There are many occasions in SPSS when users want to refer to multiple variables (e.g., to correlate a collection of variables). We could list all of the variables individually, but SPSS provides several shortcuts. In SPSS, the keyword TO can be used to refer to all variables from the first variable named to the last variable named. For example, if the file contains the 7

variables a b c x y z and d, a TO x would include variables a b c and x, c TO y would include variables c x and y, and a b y TO d would include variables a b y z and d. Note in the last example that individually named variables can be combined with the TO method. Most of the commands of SPSS accept such lists of variables as options. In general descriptions of SPSS commands, these lists will be represented by the lowercase pseudo-word "varlist."

A similar feature can be used to enter data. To enter 30 scores for each subject into v01 to v30, for example, the *DATA LIST* command could refer simply to *v01 TO v30*; SPSS would generate the intermediate values.

Occasionally, it is useful to make use of a temporary (or scratch) variable. Scratch variables can be useful, for example, when you want to transform or compute new variables (see later examples). To create a scratch variable, put a number symbol (#) in front of the variable name (e.g., #TEMP, #V1). Scratch variables are not saved for analysis by subsequent statistical procedures.

Creation and Modification of Datasets

Before SPSS can analyze data (e.g., calculate means and correlations), variables must be named and corresponding data read by the program. Sometimes, researchers also want to modify the data before analyzing it, and perform a variety of operations on the data set. This section describes the general conventions and commands used by SPSS to perform such functions.

Box S.1 presents several SPSS commands used for the creation of datasets, and for other nonstatistical purposes. The DATA LIST command states the format of the data and the variable names. Most studies can be accommodated by the FREE format option for the DATA LIST command (i.e., observations are listed sequentially and separated by spaces or commas, optionally with each subject's data on a new line). For other situations (e.g., no spaces between data), other formats are available in SPSS. Following a /, the names of the variables are listed in the order that the values appear in the following data.

```
DATA LIST FREE /varlist [(A#)].
  Note: v# TO v# creates v01 v02 v03 ...

BEGIN DATA
lines of data separated by blanks or commas
for FREE format
END DATA.

LIST [varlist]
  [/CASES FROM start TO end BY step].
```

Box S.1 - SPSS Commands for Dataset Operations.

The actual values to be read into the data set appear after a *BEGIN DATA* command, and are followed by the *END DATA* command. One simple and orderly way to arrange the data is with each subject's data on a separate row, with the individual observations aligned in columns with one or more spaces between different values. Remember to have as many variable names as there are observations for each case. If there are 2 scores for each of 6 cases, for example, there would be 2 variables listed in the *DATA LIST* command and 12 scores in the *BEGIN DATA ... END DATA* command.

Once entered (or generated by procedures explained later), data can be operated on and modified in various ways. The *LIST* command prints out the data set, and is usually a good operation to perform early on in the analysis for verification of data entry.

Box S.2 shows some SPSS commands to modify existing variables or create new variables. The RECODE command changes the values of the variable. For example, the numbers 3 and 4 for a variable called group might be recoded into a new variable called group2 as 1 and 2 by the command: *RECODE group (3=1) (4=2) INTO group2*.

```
RECODE varlist (vallist = value) (vallist = value)
...
      [/INTO varlist] [/varlist ...].
   Note - Keywords: LO HI THRU ELSE MISSING SYSMIS

COMPUTE variable = expression.
IF (condition) variable = expression.
   Notes
   Condition Operators: = > < <= >= <> AND OR NOT
   Expression Operators: + - * / ** RND() TRUNC()
       SQRT() EXP() LG10()  SUM() MEAN()

VARIABLE LABELS varname 'label' [/varname...].

VALUE LABELS varlist value 'label' value 'label'...
      [/varlist..].

MISSING VALUES varlist (valuelist) [/varlist...].

RENAME VARIABLES (varname = varname) ...  .

TITLE 'text of title in quotes'.
SUBTITLE ' text of subtitle in quotes'.

COMMENT or * text of comment.
commands   /* comment on same line as commands [*/]
```

Box S.2 - SPSS Commands for Dataset Operations.

The COMPUTE command permits researchers to calculate values based on available data and store the results in a new (or existing) variable. For example, we might use SPSS to score a test with 5 questions by computing the sum of scores on individual items: *COMP mark = q1+q2+q3+q4+q5*, or *COMP mark = SUM (q1 TO q5)*. IF performs similar

calculations, but contingent on a particular condition; omputation are performed only for cases that meet the specified condition and not for other cases. For example, *IF (group = 2) mark = mark + 5*, would add 5 to the mark variable only for cases with the group variable equal to 2. The mark for cases with group = 1, 3, or any value other than 2 would not be changed.

The *VARIABLE LABELS* command allows the user to assign a longer and more descriptive name to the variable (e.g., VAR LABELS group 'Experimental Condition'). The *VALUE LABELS* command permits users to assign labels to different values of variables (e.g., VALUE LABELS gender 1 'FEMALE' 2 'MALE'). The MISSING VALUES command tells SPSS to treat certain values as missing. For example, if someone didn't answer the question about GENDER, the researcher could include a score of 9 and tell SPSS that 9 represents a missing value. Missing values are ignored in statistical analyses (unless the user instructs SPSS to include them). The RENAME VARIABLES command permits users to assign a new name to existing variables.

Other commands shown in Box S.2 perform a variety of functions. COMMENT or * permits the user to include descriptive comments in the job and are ignored by SPSS. To add a comment after column 1, put /* in front of it (and */ behind it, if more SPSS commands follow the comment on the same line). The TITLE and SUBTITLE commands are used to label jobs and specific analyses so that the output is more easily interpreted.

Using SPSS as a "Calculator"

Understanding statistics benefits from actually carrying out the operations to produce final statistics and intermediate quantities. It is possible to use some of the features of SPSS just mentioned to help with these calculations. Imagine, for example, that certain statistics and explanations for our sample dataset were based on the sum of y, the sum of y^2, the sum of y minus a certain value (say 3.50), and the sum of the preceding deviation scores

```
COMPUTE y2 = y**2.
COMPUTE y3 = y - 3.50.
COMPUTE y4 = y3**2.
LIST.

   y       y2       y3       y4
1.00     1.00    -2.50     6.25
2.00     4.00    -1.50     2.25
3.00     9.00     -.50      .25
4.00    16.00      .50      .25
5.00    25.00     1.50     2.25
6.00    36.00     2.50     6.25

DESCR y y2 y3 y4 /STAT = SUM.

                N Sum
 y              6 21.00
 y2             6 91.00
 y3             6   .00
 y4             6 17.50
```

Box S.3 - SPSS as a Sophisticated Calculator.

squared, that is, sum of $(y - 3.50)^2$.

Box S.3 shows how SPSS can be used to perform these calculations. We create three new variables using COMPUTE: y2 is y squared, y3 is y-3.50, and y4 is y3 squared. LIST shows the resulting values have been added to the dataset. We then use DESCRIPTIVES to obtain the sums for our four variables: y, y2, y3, and y4. Anytime that you find yourself carrying out some operation repeatedly (e.g., subtracting 3.5, squaring numbers), consider whether it might not be easier to do perform the operations using SPSS as a calculator.

Statistical Analyses

SPSS has a large repertoire of data analytic techniques, some of which are discussed in this book. The general format of SPSS commands is: (a) an SPSS keyword that invokes the command (e.g., *DATA LIST, TTEST, CORRELATION*), (b) options specified by additional keywords (e.g., *FREE, GROUPS =, VARIABLES =*), (c) the user provides variable information for some of the commands (e.g., variable names in the *DATA LIST* command, *GROUPS* and *VARIABLE* variables for the *TTEST* command), and (d) a single backslash (/) often separates SPSS options.

The format in which commands will be presented is as follows. Uppercase words are SPSS keywords and lowercase words represent specific values entered by the user. To illustrate, the command *CORRELATIONS a b c /STATISTICS = ALL* invokes the *CORRELATION* procedure, and the */STATISTICS = ALL* instructs SPSS to compute and print the default statistics available for the correlation command. The letters a, b, and c are the names of the variables that the researcher wants correlated, and would usually have been introduced on the *DATA LIST* command or entered directly into the Data Editor. The list of variables is often abbreviated as "varlist" in descriptions of SPSS commands (e.g., *CORR varlist*). Most SPSS commands and parameters can be abbreviated; for example, the preceding example could have been typed as *CORR a b c /STAT = ALL*. The uppercase/lowercase convention is simply used here to disinguish SPSS keywords and user data. Commands do not need be entered specifically in upper or lower case characters; SPSS ignores the difference.

CONCLUSIONS

This brief introduction to SPSS should get you started. You will learn more about SPSS in other chapters, and can also obtain additional information from the Help option on the menu. Help provides access to an on-line manual that describes the syntax for various SPSS commands and describes briefly the nature of the analyses. Depending on how SPSS has been configured, you may have access to other useful help functions (e.g., a Tutorial, a Statistical "coach"). Spend some time exploring the Help menu and its options to learn what useful information is available.