Chapter 1

The Nature of SPSS

1.1 GETTING STARTED WITH SPSS FOR WINDOWS

Windows

SPSS for Windows is a versatile computer package that will perform a wide variety of statistical procedures. When using SPSS, you will encounter several types of windows. The window with which you are working at any given time is called the *active* window. Four types of windows are:

Data Editor Window. This window shows the contents of the current data file. A blank data editor window automatically opens when you start SPSS for Windows; only one data window can be open at a time. From this window, you may create new data files or modify existing ones.

Output Viewer Window. This window displays the results of any statistical procedures you run, such as descriptive statistics or frequency distributions. All tables and charts are also displayed in this window. The viewer window automatically opens when you create output.

Chart Editor Window. In this window, you can modify charts and plots. For instance, you can rotate axes, change the colors of charts, select different fonts, and rotate three-dimensional scatter plots.

Syntax Editor Window. You will use this window if you wish to use SPSS syntax to run commands instead of clicking on the pull-down menus. An advantage to this method is that it allows you to perform special features of SPSS that are not available through dialog boxes. Syntax is also an excellent way to keep a record of your analyses.

To start an SPSS session, select SPSS from the programs submenu on the Windows Start menu. Figure 1.1 shows what the screen will look like when SPSS for Windows first opens.

The Main Menu

SPSS for Windows is a menu-driven program. Most functions are performed by selecting an option from one of the menus. We refer to these menus as "pull down" menus since an entire menu of options appears when one is selected. The main menu bar is where most functions begin, and is located at the top of the window (see Fig. 1.1). Any menu may be activated by simply clicking on the desired menu item, or using the Alt-letter keystroke (each menu uses the first letter in the menu word). For example, to activate the file menu, either click the mouse on **File** or use the keyboard with **Alt-F**. The main menu bar lists 10 menus:

File. This menu is used to create new files, open existing files, read files that have been created by other software (e.g., spreadsheets or databases), and print files.

Edit. This menu is used to modify or copy text from output or syntax windows.

View. This menu allows you to change the appearance of your screen. You can, for instance, change fonts, customize toolbars, and display data using their value labels.

Data. Use this menu to make temporary changes in SPSS data files, such as merging files, transposing variables and cases, and selecting subsets of cases for analyses. Changes are not permanent unless you explicitly save the changes.

Transform. The transform menu makes changes to selected variables in the data file and computes new variables based on values of existing variables. Transformations are not permanent unless you explicitly save the changes.

Analyze. Use this menu to select a statistical procedure to be performed such as descriptive statistics, correlations, analysis of variance, and cross-tabulations.

Graphs. This menu is used to create bar charts, pie charts, histograms, and scatter plots. Some procedures under the Analyze menu also generate graphs.

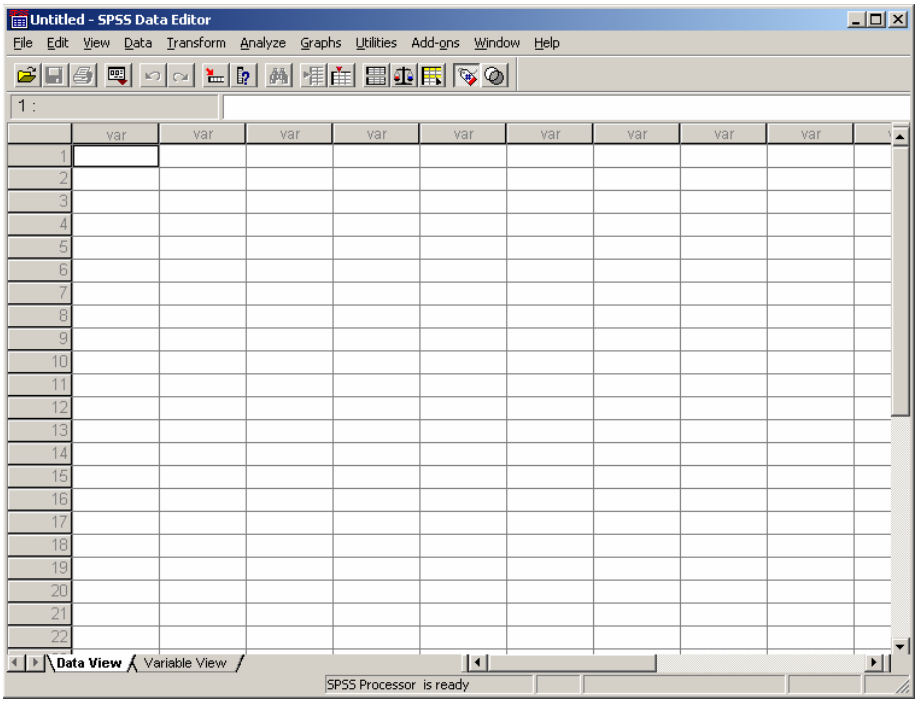

Figure 1.1 SPSS Data Editor

Utilities. This menu is used to change fonts, display information on the contents of SPSS data files, or open an index of SPSS commands.

Window. Use the window menu to arrange, select, and control the attributes of the SPSS windows.

Help. This menu opens a Microsoft Help window containing information on how to use many SPSS features.

1.2 MANAGING DATA AND FILES

Entering and selecting data files in SPSS for Windows is quite easy. We will demonstrate how to enter raw data from scratch and how to open existing data files.

Entering Your Own Data

Raw data may be entered in SPSS by using the SPSS data editor. (ASCII data may also be entered with another editor, which are then read by SPSS using the

Text Import Wizard as described below.) The SPSS editor looks like a spreadsheet or grid and is automatically opened each time you start an SPSS session. The editor is used to enter, edit, and view the contents of your data file. If you are opening an existing data file, the data will appear in the editor and you may then use the editor to change the data or add or delete cases or variables. If you are starting from scratch and wish to enter data, the data editor will be empty when it is first opened.

The data editor is a rectangle defined by rows and columns. Each cell represents a particular row-by-column intersection (e.g., row 1, column 3). All data files in the data editor have a standard format. Each row of the data editor represents a case (e.g., subject #1 or John Doe). Each column in the data editor represents a variable (e.g., heart rate or sex). Cells in the editor may not be empty. That is, if the variable is numeric and there is no valid value, the cell is represented by a "system-missing" value and a period appears in the cell. If the variable is a string variable, a blank cell is considered to be valid. (See Section 1.5 for further information on the treatment of missing values.)

To begin entering data in the data editor, follow these steps:

1. Click on **File** from the menu bar.
2. Click on **New** and then **Data** from the file pull-down menu.
3. Click on the cell in which you wish to enter data (or use the arrow keys to highlight the cell). Begin at the uppermost left cell in the rectangle. This is row 1, column 1. Once you have clicked on that cell, a darkened border will appear around the cell; this tells you that this is the cell you have selected.
4. Type in the value you wish to appear in that cell and then press Enter. You should notice that the value you type will appear at the top of the data editor window and in the cell. Notice that entering a value in this first column and pressing Enter automatically creates a variable with the default name VAR00001, which appears at the top of the column. Later we will demonstrate how to specify original names and alternate formats for variables. As an example, suppose that you are recording ages for 25 people. If the age of the first person is 18, enter 18 in the first cell.
5. Type in another value for the second case. This cell is directly below the previous cell. This location is row 2, column 1. Again, you will see the value at the top of the data editor and in the cell. Suppose the age of the second person was 22, enter 22 in row 2 column 1.
6. Repeat this process until you have entered all of the data you wish for column 1 (values for all cases on variable 1).
7. When you are ready to add another variable, click on the first cell in the next column (row 1, column 2). Suppose that "shoe size" is the next variable, and the first person has size 7. Enter this value and press enter. This will automatically create a new variable and call it VAR00002.
8. Repeat this process for all values in column 2.

9. Continue this procedure until you have entered values for all cases and variables that you wish for your data file.

Once you have entered data in the data editor, you may change or delete values. To change or delete a value in a cell, simply click on the cell you wish to alter. You will notice that a dark border appears around the selected cell, and the value in the cell appears at the top of the data editor. If you are changing the value, simply type the new value and press enter. You should see the new value replace the old value in the cell.

Adding Cases and Variables

To insert a new case (row) between cases that already exist in your data file:

1. Point the mouse arrow and click on the row number *below* the row where you wish to enter the new case. The row should be highlighted in black.
2. Click on **Data** on the menu bar.
3. Click on **Insert Cases** from the pull-down menu.

A new row is now inserted and you may begin entering data in the cells. Notice that before you enter your values, all of the cells have system-missing values (represented by a period).

To insert a new variable (column) between existing variables:

1. Click on the column variable name that is to the *right* of the position where you wish to enter a new variable. The column should be highlighted in black.
2. Click on **Data** on the menu bar.
3. Click on **Insert Variable** from the pull-down menu.

A new variable (column) is now inserted and you may begin entering data in the cells.

Deleting Cases and Variables

To delete a case:

1. Click on the case number that you wish to delete.
2. Click on **Edit** from the menu bar.
3. Click on **Clear**.

The selected case will be deleted and the rows below will shift upward.

To delete a variable:

1. Click on the variable name that you wish to delete.
2. Click on **Edit** from the menu bar.
3. Click on **Clear**.

The selected variable will be deleted and all variables to the right of the deleted variable will shift to the left. Deleting variables can also be accomplished using SPSS syntax (see Section 1.6) with the Drop and Keep subcommands.

Defining Variables

By default, SPSS assigns variable names and formats to all variables in the SPSS data file. By default, variables are named VAR##### (prefix VAR followed by five digits) and all values are valid (blanks are assigned system-missing values). Most of the time, however, you will want to customize your data file. For example, you may want to give your variables more meaningful names, provide labels for specific values, change the variable formats, and assign specific values to be regarded as "missing." To do any or all of these:

1. First, make sure that your data file window is the active window and click on the variable name that you wish to change.
2. Click on the **Variable View tab** or else double-click on the variable name in the data editor.
3. Type the name of the variable in the Name column. Variable names have to be unique, begin with a letter, and cannot contain blank spaces.
4. If you wish to change the type or format of a variable, click the button in the Type cell to open the Variable Type dialog box. By default, all variables are numeric, but you may work with other types such as names, dates, and other non-numeric data. Suppose you have a variable that contains letters (e.g., student names). This is known as a string variable and you would indicate this by clicking on **String** in the Variable Type dialog box and then clicking on **OK**.
5. Suppose you have a variable representing average cost of groceries per person that was entered to the nearest cent (e.g., 32.24) and you want to change this format so that the average cost is displayed as a whole number (rounded to the nearest dollar, e.g., 32). To change the format of the numeric variable, click in the **Width** box. The number in this box tells you the total number of columns that the variable occupies in the data file (including one column for decimal places, plus, or minus signs). For example, 8 indicates that the variable is 8 columns wide. Use the arrows to adjust the variable's column width. If you wish to change the number of decimal places, click in the **Decimals** box. The number in this box tells you how many numbers appear after the decimal place. For example, the number

32.24 would have a "width" of 5 and a 2 in the "decimal places" box. The number 32 would have a width of 2 and a 0 in the decimal places box. Use the arrows to adjust the number of decimal places.

6. If one of your variables is categorical, you can assign numbers to represent the categories of the variable. For example, the variable sex will have 2 categories: male and female. Males may have the assigned value "1" and "2" represents females. It is useful to have descriptive labels assigned to the values of 1 and 2 so that it is easy to see which number represents which category in your output files.

To assign value labels to the variable, click the button in the Values cell to open the Value Labels dialog box. Type the number representing the first category (e.g., 1) in the Value box. Type the corresponding value label (e.g., male) in the Value Label box. Click on the **Add** button. Go back to the Value box and type in the next value (e.g., 2). Type the value label for this value in the Value Label box (e.g., female), and click on **Add**. Note that each time you click Add, you will see the value and its' corresponding label appear in the window to the right of the Add button. When you have added all of the values and labels, click on **OK**.

7. If there are specific values that you would like to be treated as missing values, click on **Missing** to open the Missing Values dialog box. Click on **Discrete Missing Values** to tell SPSS that you have specific values that are considered to be missing. Type the value(s) in the boxes (you may have up to three values). If you have more than three missing values, click on **Range plus one optional discrete missing value** and enter the lower and upper bounds of the discrete variable. Click **OK** when you have entered in all of your missing values.

Opening Data Files

SPSS for Windows can read different types of data already entered into computer files. The file type we will use in this manual is the SPSS data file. These files are easily identified because (by default) each file name is followed by an ".sav" extension. SPSS data files are unique because they contain the actual data as well as information about the data such as variable names, formats, and column locations. These files are written in a special code that is read and interpreted by the SPSS program. If you try to read these data files with software other than SPSS, the file will look like lines of secret code and will not make sense to you. However, they make a great deal of sense to SPSS, and this is why reading them with SPSS is so easy. If you would like to look at the information contained in an SPSS data file (that is currently open), you can do this by clicking on **File** in the menu bar, and then choose **Display Data File Information**

and then **Working File**.

SPSS for Windows can also read raw data that are in simple text files in standard ASCII format. Text files are usually identified by a ".dat" or ".txt" extension. These are data files that just contain ordinary numbers (or letters). There is no additional information contained in the file such as variable locations, formats, labels, missing values, etc. (SPSS .sav data files do contain this additional information). You can read text files with many different software programs, including WordPad. SPSS can read text data files that are formatted as fixed or tab-delimited.

The SPSS Data Editor is designed to read a variety of formats in addition to SPSS data files and ASCII text files. For example, spreadsheet files created with Lotus 1-2-3 and Excel, database files created with dBASE and SQL formats, and SYSTAT data files.

Reading SPSS Data Files

We will illustrate how to read an existing SPSS data file. The reader may follow along using the data accompanying this guide.

To open a data file:

1. Click on **File** from the menu bar.
2. Click on **Open** on the file pull-down menu.
3. Click on **Data** on the open pull-down menu. This opens the Open File dialog box as shown in Figure 1.2.
4. Choose the correct directory from the **Look in:** box at the top of the screen.
5. Point the arrow to the data file you wish to open and click on it. By default, all SPSS data files (*.sav) in the current directory will be displayed in the list. If your data file is not visible in the file name box, use the left and right arrows to scroll through the files until you locate your desired file. Note that all of the SPSS data files have the .sav extension, and this is designated in the Files of type window. Before you open a data file, make certain that the file type is correct. If you are reading SPSS data files and the file type box does not read "SPSS (*.sav)," you must scroll through the file types and select that type. For example, to open the file called "football.sav," highlight the name of this file by clicking on it with the mouse button.
6. Click on **Open**. You should now see the contents of the data file displayed in the Data Editor window. The "football.sav" data file contains two variables, "height" and "weight," for 56 football players from Stanford University. The variable names are displayed at the top of the Data Editor; each column contains one variable. The rows in the data file are the cases; in this data file there are 56 cases.

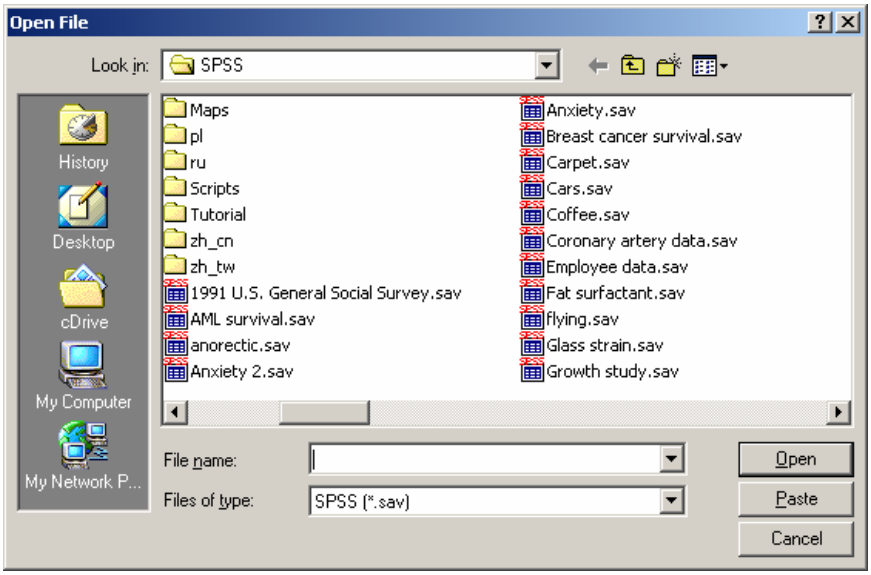

Figure 1.2 Open File Dialog Box

Note: Most of the examples in the following chapters use the SPSS data files that are provided with this manual. Unless you are required to enter data on your own into a new file, all procedures assume that you have opened the SPSS data file before beginning any computations or analyses.

Reading Data Files in Text and Other Formats

To read a text data file, begin at the main menu bar in the Data Editor window:

1. Click on **File.**
2. Click on **Read Text Data**.
3. Select the appropriate file from the Open file dialog box and click **Open**.
4. Follow the steps in the **Text Import Wizard** to read the data file. You will have to answer questions about type of data, arrangement of data, number of cases to import, and missing values. Use the Help button of the Text Import Wizard for more detailed information.

To open data from a file such as an Excel spreadsheet, begin at the Data Editor window:

1. Click on **File.**
2. Click on **Open** and then click on **Data**.

3. Select the file format from the drop-down list of file types in the **Files of type:** box.
4. Choose the appropriate directory and file.
5. Click on **Open**.

Excel, Lotus, and SYLK variable names are read from the file and appear in the first row of the spreadsheet. If the spreadsheet does not contain variable names, SPSS provides default names using column letters.

Saving Data Files

Unless you save your files, all of your data and changes will be lost when you leave the SPSS session. To save a file, first make the Data Editor the active window. Then:

1. Click on **File** from the menu.
2. Select **Save** from the list of options in the File pull-down menu.
3. Select the appropriate directory in the **Save in:** box. Type the name of your file in the **File name** box. Notice that the default file type is set for SPSS format as indicated by the ".sav" extension.
4. Click on **Save**.

By default, this will save the data file as an SPSS data file. If you were working with a previously existing data file, the old file will be overwritten by the modified data file. To save the file with a different name, select **Save As …** from the File pull-down menu. Note that it is always recommended that you preserve your original data in a separate file in case you ever need to return to it.

If you wish to save the data file in a format other than SPSS (e.g., Lotus, Excel, dBASE, fixed-format ASCII text):

1. Click on **File** from the menu.
2. Select **Save As** from the list of options in the File pull-down menu.
3. Select the appropriate directory in the Save in: box. Type the name of your file in the **File name** box.
4. Choose the appropriate file type in the **Save as type:** box.
5. Click on **Save**.

1.3 TRANSFORMING VARIABLES AND DATA FILES

At times, you may need to alter or transform the data in your data file to allow you to perform the calculations you require. There are many ways in which you

can transform data. This section discusses three commonly used techniques: computing new variables, recoding variables, and selecting subsets of cases.

Computing New Variables

There may be occasions when you need to compute new variables that combine or alter existing variables in your data file. For instance, your data file may contain daytime and nighttime sleeping hours for a sample of infants, but you are interested in examining total sleep hours (i.e., the sum of the separate daytime and nighttime hours).

To create a new variable:

1. Click on **Transform** from the menu bar.
2. Click on **Compute** from the pull-down menu. This opens the Compute Variable dialog box (see Fig. 1.3).
3. Enter the name of the new variable (in the above illustration, total) in the Target Variable box. (You also have the option to describe the nature and format of the new variable by clicking on the **Type & Label** box.)
4. You will then need to perform a series of steps to construct an expression used to compute your new variable. In this illustration, you would first select the daytime variable ("daysleep") from the variable list box on the left-hand side of the dialog box and move it to the Numeric Expression box using the **right directional arrow.**
5. Then click on the "+" from the calculator pad. You will notice that a plus sign is placed in the Numeric Expression box after the word daytime.
6. Complete the expression by selecting the nighttime variable ("nightsleep") and moving it to the Numeric Expression box, following the instructions in step (4) above.
7. When you have completed the expression, click on **OK** to close the Compute Variable dialog box. Your new variable will be added to the end of your data file.

In addition to simple algebraic functions on the calculator pad (+, -, x, ÷), there are many other arithmetic functions such as absolute value, truncate, round, square root, and statistical functions including sum, mean, minimum, and maximum. These are displayed in the Function group box to the right of the calculator pad. First, select a procedure in the Function group window, and then select the specific function in the Functions and Specific Variables window.

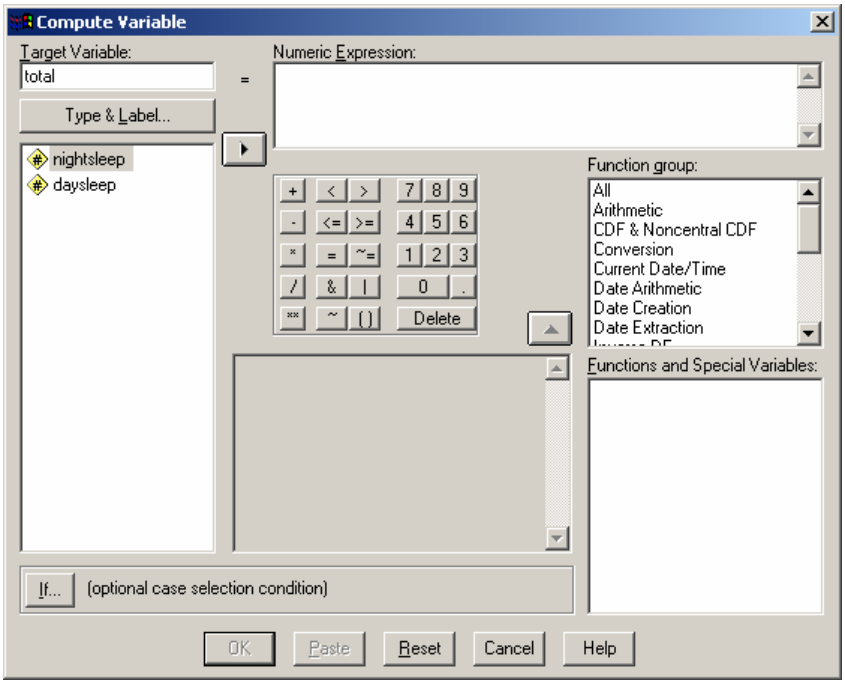

Figure 1.3 Compute Variable Dialog Box

Recoding Variables

Recoding variables is often a useful technique. There are several instances in which you may need to modify the values in your original data file. You can recode either categorical or numeric variables. For instance, you may have a data file containing information about Olympic athletes such as number of medals won, years of training, and type of sport. A "sport" variable, with a value of 1 to 100, indicates the type of event (e.g., bobsled, diving, etc). Suppose that you are interested in examining the medal distribution based on whether athletes competed in the summer or winter Olympics. For this, you would need to recode the sport variable from 100 categories to two.

Or, you may want to categorize a discrete or continuous numeric variable into a limited number of groups. For example, your data file may contain the number of years of training and you wish to group them into three categories: 1–5 years, 6–10 years, and more than 10 years.

You have two options available for recoding variables. You may recode values into the same variable, which eliminates all record of the original values. This is a useful function for correcting obvious data errors, changing system

missing values into a valid value for profiling item nonrespondents, or collapsing a number of values when only a few cases responded in a particular way and a meaningful assessment of these few cases cannot be conducted. You also have the option to create a new variable containing the recoded values. This preserves the values of the original variable. If you think that there may be a reason that you would need to have record of the original values, you should select the second option.

Recoding into the Same Variable

To recode into the same variable:

1. Click on **Transform** from the main menu.
2. Click on **Recode** from the pull-down menu.
3. Click on **Into Same Variables** to open the Recode into Same Variable dialog box.
4. Select the name of the variable to be recoded, and move it to the Variables box with the **right arrow button**.
5. Click on **Old and New Values**. This opens the Old and New Variables dialog box (see Fig. 1.4).
6. For each value (or range of values) you want to recode, you must indicate the old value and the new value, and click on **Add**. The recode expression will appear in the **Old --> New** box.

Old values are those before recoding — the values that exist in the original variables. There are several alternatives for recoding old values, including the Value and the Range options discussed below.

Value Option. You may use the Value option for cases in which you want to collapse existing categorical variables such as the "sport" illustration above. In your original data file, you have an existing value (1 through 100) for each type of sport, but you need to recode these values into the values 1 and 2 — 1 representing summer and 2 representing winter. For example:

Original		**Recoded**	
Category	**Value**	**Category**	**Value**
Diving	1	Summer	1
Gymnastics	2	Summer	1
Skating	3	Winter	2
Bobsled	4	Winter	2

The steps to perform this recoding are straightforward:

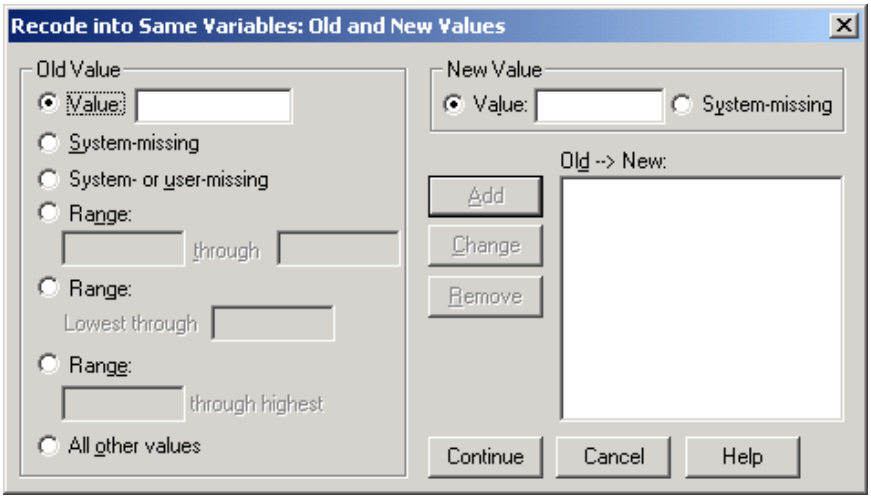

Figure 1.4 Recode into Same Variables: Old and New Values Dialog Box

1. Type **1** in the **Value** box of the **Old Value** section, indicating the existing value for Sport 1 (Diving).
2. Type a **1** in the **Value** box of the **New Value** section, indicating that it is to be recoded as season 1 (Summer).
3. Click on **Add**. You will notice that the expression 1 --> 1 will appear in the **Old -->New** box.

Follow the same procedure for the rest of the values of the original variable. For example, you would recode the old value of 2 to the new value of 1, and click on **Add**. Note that because Diving was coded 1 both before and after re-coding, it was not necessary to include it in the recode procedure. Doing so, however, may assist you in making sure that you have included all values to be recoded.

Range Option. You may also recode variables using the Range option. This is most useful for numerical variables. The procedure is similar to that discussed above. To recode the years of training in this example, the existing values would be recoded as follows:

Original **Range of Values**	**Recoded** **Value**
Lowest through 5	1
6 through 10	2
11 through highest	3

Instead of choosing the value option in the old value section, you may use the range option as follows:

1. Type **5** in the **Range: Lowest through ___** box ; the middle range option box.
2. Type **1** in the **Value** box under the **New Value** section.
3. Click on **Add**. The expression will appear in the **Old --> New** box.
4. Type **6** and **10** in the two boxes of the first range option: **Range: ___ through ___**.
5. Type **2** in the **Value** box under the **New Value** section.
6. Click on **Add**.
7. Type **11** in the **Range: ___ through highest** box; the bottom range option of the Old Value section.
8. Type **3** in the Value box under the **New Value** section.
9. Click on **Add**.

When you have indicated all the recode instructions, using either the Value or Range method, click on **Continue** to close the Recode Into Same Variables: Old and New Values dialog box. Click on **OK** to close the Recode Into Same Variables dialog box. While SPSS performs the transformation, the message "Running Execute" appears at the bottom of the application window. The "SPSS Processor is Ready" message appears when transformations are complete.

Recoding into Different Variables

The procedure for recoding into a different variable is very similar to that for recoding into the same variable:

1. Click on **Transform** from the main menu.
2. Click on **Recode** from the pull-down menu.
3. Click on **Into Different Variables** to open the Recode into Different Variable dialog box.
4. Select the name of the variable to be recoded from the variables list, and move it to the Input Variable --> Output Variable box with the **right arrow button**.
5. Type in the name of the new variable you wish to create in the Output Variable box. If you wish, you may also type a label for the variable.
6. Click on **Change**, and the new variable name will appear linked to the original variable in the Input Variable --> Output Variable box.
7. Click on **Old and New Values**. This opens the Old and New Variables dialog box.

8. The procedure for identifying old and new values is the same as that discussed in the Recoding into the Same Variable subsection, with one exception. Because you are creating a new variable, you must indicate new values for *all* of the old values, even if the value does not change. (This is optional when recoding to the same variable.) Because this step is mandatory, SPSS provides a **Copy Old Value(s)** option in the New Value box.
9. When you have indicated all the recode instructions, click on **Continue** to close the Recode Into Different Variables: Old and New Values dialog box.
10. Click on **OK** to close the Recode Into Different Variables dialog box. While SPSS performs the transformation, the message "Running Execute" appears at the bottom of the Application Window. The "SPSS Processor is Ready" message appears when transformations are complete, and a new variable appears in the data editor window. The new variable is added to your data file in the last column displayed in the Data Editor window.

Selecting Cases

There may be occasions when you need to select a subset of cases from your data file for a particular analysis. You may, for instance, have a data file containing height and weight for 200 individuals, but you need to know the average height of individuals over 120 pounds. Or, you may simply wish to select a random sample of cases from a very large data file.

To select subset of cases:

1. Click on **Data** from the main menu.
2. Click on **Select Cases** from the pull-down menu. This opens the Select Cases dialog box (see Fig. 1.5).

There are several ways in which you can select a subset from a larger data file. We will discuss the If Condition and Random Sample methods.

If Condition

In the height and weight example given above, you would need to:

1. Select the **If condition is satisfied** option and click on **If** to open the Select Cases: If dialog box.
2. Select the weight variable from the variable list box, and move it into the box above the calculator pad with the **right arrow button**.
3. Using the calculator pad, click on >. This sign will appear in the upper box.

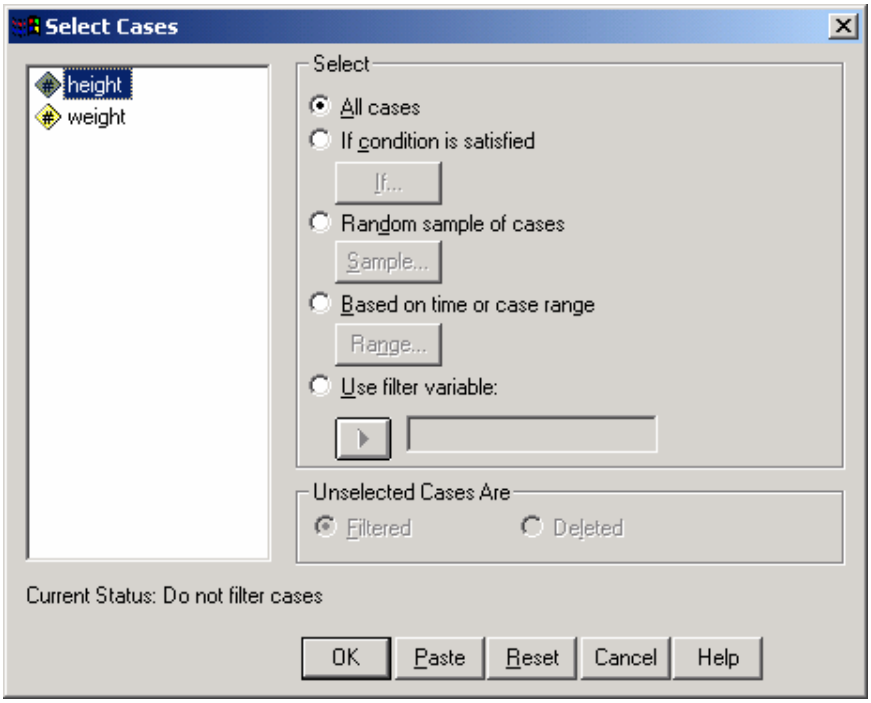

Figure 1.5 Select Cases Dialog Box

4. Using the number pad, Click on **1**, then **2**, then **0**, to create the expression "weight > 120" in the upper right-hand box.
5. Click on **Continue** to close the Select Cases: If dialog box.

Random Sample

Another method for selecting subcases is to choose a random sample from your data file:

1. From the Select Cases dialog box (see Fig. 1.5, above) select the **Random sample of cases** option and click on **Sample** to open the Select Cases: Random Sample dialog box.
2. Type either a percentage or a precise number of cases in the appropriate box.
3. Click on **Continue** to close the Select Cases: Random Sample dialog box.

You should now be back at the Select Cases dialog box. Click on **OK** to

close this dialog box. The message "Running Execute" will appear as SPSS processes your instructions. SPSS creates a new "filter" variable with the value 1 for all cases selected and 0 for all cases not selected. There is also a "Filter On" message at the bottom of the application window to remind you that your subsequent analyses will be performed only on the designated subset of data. Furthermore, there are diagonal lines through the row line numbers of the Data Editor window for all cases not selected. SPSS uses the filter variable to determine which cases to include in subsequent analyses. Unselected cases remain in the data file but are not included in subsequent analyses. You can use the unselected cases by turning the filter off. To turn the filter off, go back into the Select Cases dialog box and click on **All Cases** in the Select box, and then click on **OK**.

1.4 MISSING VALUES

In many situations, data files do not have complete data on all variables, that is, there are missing values. You need to inform SPSS when you have missing values so that all computations are performed correctly. With SPSS, there are two forms of missing values: system-missing and user-defined missing.

System-missing values are those that SPSS automatically treats as missing (without the user having to explicitly inform SPSS). The most common form of this type of value is when there is a "blank" in the data file. For example, the data value for a person is missing if the information was not provided. A period is displayed in the data file cell that does not have a value. When SPSS reads this variable, it will read a blank, and thus treat the value as though it is missing. Any further computations involving this variable will proceed without the missing information. For instance, suppose you wish to calculate the average amount of daily water consumption for a sample of 20 adults, but you only have data entered for 19 people. SPSS will read the "valid" values for the 19 adults, ignore the missing value, and compute the average based on the 19 individuals.

User-defined missing values are those that the user specifically informs SPSS to treat as missing. Rather than leaving a blank in the data file, numbers are often entered that are meant to represent missing data. For example, suppose you have an age variable that ranges from 1 through 85. You could use the number 99 to represent those individuals who were missing information on age. (You could not use any numbers from 1 to 85 since these are all valid values.)

In this example, you need to inform SPSS that the value 99 is to be treated as a missing value, otherwise it will treat is as valid. This is explained in Section 1.2, but in brief, you need to do the following: switch to Variable view in the data editor and click the "..." button in the **Missing** column. Enter 99 in one of the **Discrete Missing Values** boxes and click on **OK**. When SPSS reads this

variable, it will then treat 99 as a missing value and not include it in any computations involving the "age" variable. User-missing values will look like valid values in the data editor, but are internally flagged as "missing" in SPSS data files, and labeled as missing in the output for some procedures.

Most SPSS computations will display the valid number of cases in the output. This is the number of cases that were not system-missing and/or user-defined missing; these cases were used in the computations. The number of missing cases (not used in the computations) is typically displayed as well.

Analyses with Missing Data

When you have missing data, they can be treated in several ways. Missing data is a complex issue and can be problematic. If you do not specify how to handle missing data in some analyses, cases that have missing values for any of the variables named in the analysis are omitted from the analysis using pairwise or listwise deletion. For example, suppose you wish to calculate the correlations between three sets of achievement scores: math, science, and reading. Three correlations will be computed: math with science, math with reading, and science with reading. Using **pairwise deletion**, cases that do not have valid scores for both measures (e.g., math and science) will be excluded from the computation. Using **listwise deletion**, cases that do not have valid scores on all measures (i.e., math, science, and reading) will be excluded from the computations. For example, suppose you have a data file containing 1,000 cases, and all cases have a valid math score, 700 cases have a valid science score, and 400 cases have a valid reading score. Using pairwise deletion, the correlation between math and science will be based on the 700 cases have complete data on both achievement measures. Using listwise deletion, the same correlation will be based on only 400 cases because only 400 cases have complete data on all three measures. Further, any correlation you calculate with this sample would be based on 400 cases if you specify listwise deletion. As you can see, listwise deletion can greatly reduce the sample size used in your analyses. On the other hand, listwise deletion ensures that all computations are based on the same number of cases.

By default, SPSS uses either pairwise or listwise deletion depending on the procedure. For example, listwise deletion is the default for multiple regression, but pairwise deletion is the default for bivariate correlations. To designate pairwise or listwise deletion, click on the **Options** button after opening the dialog box for the appropriate analysis and then choose the method for handling missing values and click **Continue**.

In some situations, you may wish to substitute missing values with a new value to be used in the analysis rather than excluding cases from your computations. For example, in repeated measures analyses, complete data in the series is required for the analysis, so missing data are not allowed. Substituting missing values can be done with SPSS using the **Replace Missing Values** function,

which creates new variables from existing ones, replacing missing values with estimates created from one of several procedures. One of the most common procedures is to replace missing values with the mean for the entire series. Other procedures include replacing missing values with the mean or median of nearby points (as indicated by the user) or linear interpolation. Overall, missing data represents a complex problem for the data analyst and simple solutions such as replacement of missing values is generally not advised. Further, imputing missing values is not recommended for variables with a large amount of missing data.

To replace missing values with the mean for the entire series:

1. Click on **Transform** from the main menu.
2. Click on **Replace Missing Values** to open the Replace Missing Values dialog box (see Fig. 1.6).
3. Click on the variable for which you wish to substitute values and click on the **right arrow button** to move it into the New Variable(s) box. By default, a new variable name will be created by using the first six characters of the existing variable followed by and underscore and a sequential number. For example, the variable "math" would be replaced with a new variable "math_1." The new variable name will appear in the Name: field in the Name and Method box.
4. Click on **Series Mean** in the Method: box.
5. Click on **OK**.

The output shows the number of cases for which the mean math score for the series was substituted for missing math scores.

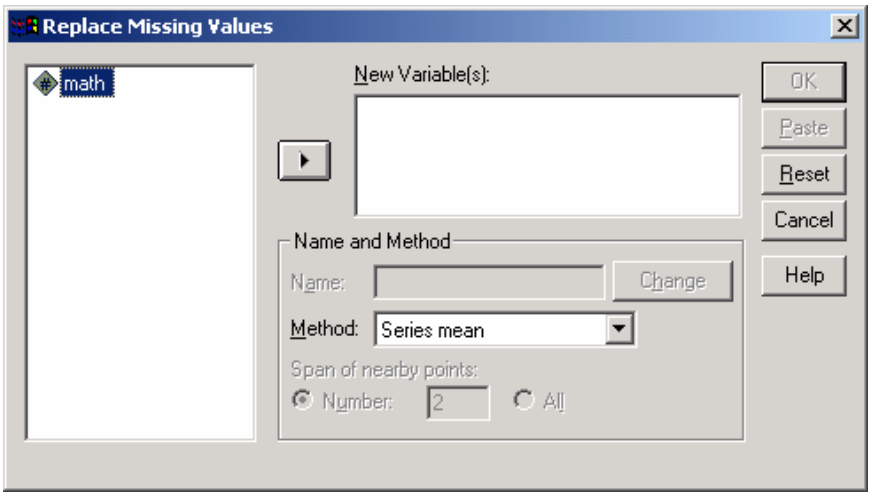

Figure 1.6 Replace Missing Values Dialog Box

1.5 EXAMINING AND PRINTING OUTPUT

After running a procedure, SPSS results are displayed in the output Viewer window (see Fig. 1.7). From this window, you can examine your results and manipulate output. The viewer is divided into two panes. An outline of the output contents is arranged on the left side. The right side contains the detailed output of your procedures such as descriptive statistics, frequency distributions, results of *t*-tests, as well as charts and plots including histograms, box-and-whisker plots, and scatter plots. Each time you direct SPSS to create a chart or graph, it displays it in the viewer. Double-click on a graph or chart if you wish to edit it. You can go directly to any part of the output by clicking on the appropriate icon in the outline in the left pane. You may also view the output by using the directional arrows buttons on the vertical scroll bar at the right edge of the window.

Printing Output

To print the contents of an Output Window:

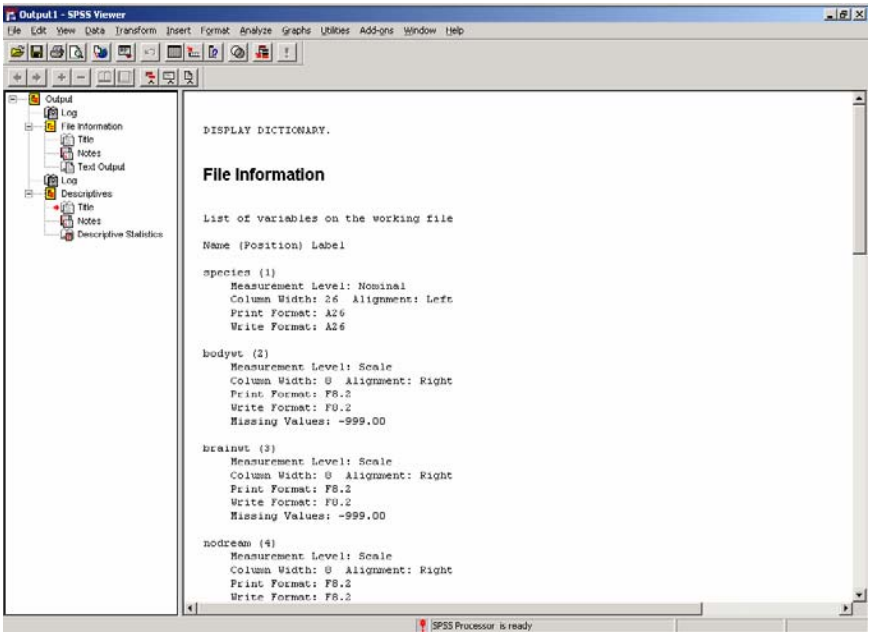

Figure 1.7 Viewer window

1. Make the viewer window the active window.
2. Click on **File** from the main menu.
3. Click on **Print** from the pull-down menu. This opens the Print dialog box.
4. If you wish to print the entire file, click on the **All Visible Output** option. If you wish to print only a selected block, click on the **Selection** option. To print only a section of the file, you need to use the "click-and-drag" method to highlight the area before opening the Print dialog box.
5. Click on **OK** to close the Print dialog box and print the file.

1.6 USING SPSS SYNTAX

As illustrated throughout this book, most SPSS procedures are conducted using the pull-down menus because they are convenient and easy to use. However, an alternative way to run SPSS procedures is through command syntax. SPSS commands are the instructions that you give the program for conducting procedures. SPSS syntax commands are typed into a command file using the SPSS syntax editor. Syntax files have the extension ".sps."

There are several reasons why command syntax is useful, such as when the user wants to: (1) have a record of the analyses conducted during a session; (2) repeat long and complex analyses; (3) review how variables were created or transformed; and (4) modify commands to run slightly different or customized statistics.

When working with syntax, the user must enter commands instructing the program what procedures to conduct. You can enter syntax by either typing or pasting syntax into the syntax editor. Because most users do not know the commands from memory, it is useful to refer to the *SPSS Syntax Reference Guide* for a complete reference to the command syntax. Help is also available by using the Help button on the toolbar in the syntax editor window. Pasting syntax commands from dialog boxes is perhaps the easiest way to construct syntax commands. Rather than typing the commands, you initiate a procedure using pull-down menus and then instruct SPSS to provide the commands and paste them into the syntax editor.

To open a new window and begin typing commands:

1. Click on **File** from the main menu.
2. Click on **New** from the pull-down menu.
3. Click on **Syntax** to open the SPSS syntax editor (see Fig. 1.8).
4. Begin typing syntax into the editor.

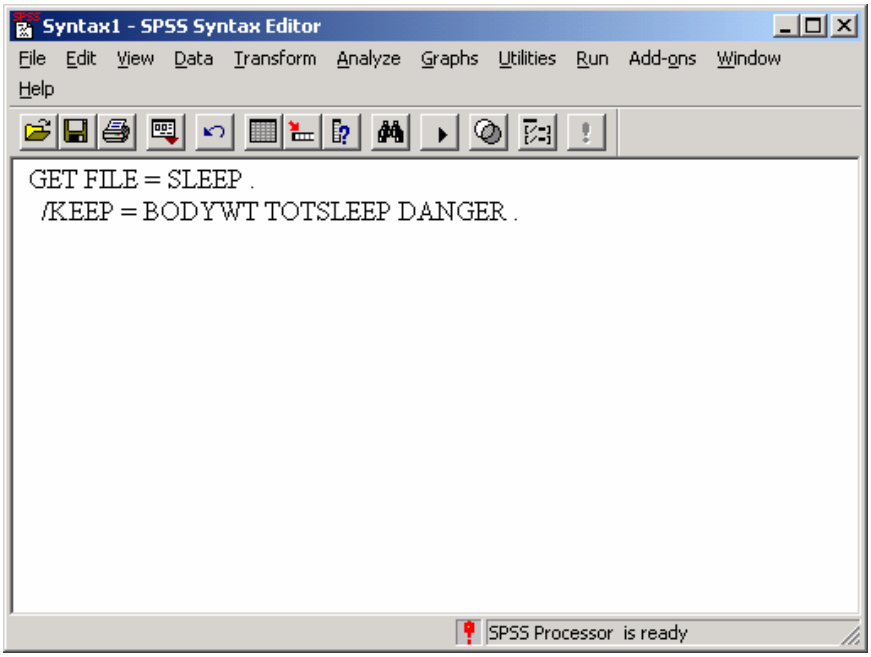

Figure 1.8 SPSS Syntax Editor

For example, suppose you want to open the sleep.sav data file, but you only want to read a subset of variables — body weight, total sleep, and danger index. The syntax command would be:

```
GET FILE = SLEEP .
 /KEEP = BODYWT TOTSLEEP DANGER .
```

You can also run a procedure by pasting syntax from a dialog box. When you use the paste button, SPSS creates the syntax commands to execute procedures requested from pull-down menus. For example, to compute a new variable (total sleep hours) as shown in Section 1.3, follow steps 1–6. Instead of clicking on **OK**, click on the **Paste** button. The compute commands will automatically be displayed in a syntax window. To run the syntax commands, click the **Right arrow button** on the toolbar.

Once you have created a syntax file, you can save it using the same procedures described in Section 1.2 of this chapter. The file can then be opened and edited for future modifications. Make sure when you open, edit, and save a syntax file that you correctly identify it with the ".sps" file type.

Chapter Exercises

1.1 Select the data file "football.sav" and without opening the file, answer the following:

- **a.** How many variables are in the file?
- **b.** What is the format for the weight variable?

1.2 Open the SPSS data file "spit.sav" and answer the following:

- **a.** Is this a text file?
- **b.** How many cases are in the file?
- **c.** How many variables are in the file?
- **d.** Are there any missing data?

1.3 Enter the age and sex for 10 students in your class into a new data file using the SPSS data editor.

- **a.** Save the data as an SPSS data file.
- **b.** Save the data as an ASCII data file.
- **c.** Once you have saved and exited the file, re-open the ASCII data file and enter a new variable named "minors" with the value of 1 for students under the age of 19 and 2 for students 19 or older. Save the file as an SPSS file.
- **d.** Re-open the SPSS data file and delete the fifth case. Was this a minor?

Part II

Descriptive Statistics