chapter 1-4

Processing Data Using SPSS

Introduction

In this chapter the basics of using SPSS for Windows will be discussed. SPSS stands for Statistical Package for the Social Sciences, and is one of the most commonly used statistical packages in university research laboratories. Written specifically for statistical analysis, it far exceeds the statistical analysis capabilities we have described for MS EXCEL.

The aim of this chapter is to show the basics of a few selected topics including opening data files, formatting variables, selecting cases for analysis and the basics of statistical analyses. The detailed use of SPSS will be a focus of each of the chapters relating to specific statistical analysis later in the text. All of the examples in this text were generated with SPSS. Earlier or later versions may present slight differences in menus or dialog boxes. If you need more detailed help than is provided here, the SPSS software includes tutorials and a useful statistics coach, which takes you through the steps of deciding which test to use on your data. In addition, there is a useful pdf file handout available on the web at:

http://ssc.utexas.edu/software/software-tutorials#SPSS

Data Entry

Data entry can be carried out directly with SPSS using the Data Editor window shown in Figure 1-4.1 in the same manner as spreadsheets. My preference, however, is to enter data in MS EXCEL and save in XLS format. Opening XLS or other formatted cells with SPSS is as simple as selecting the FILE menu OPEN DATA option, and selecting the type of data file you wish to open (Figure 1-4.1). When opening the files note that on the OPEN DATA dialog box you should select read variable names, if you had already inserted variable names in the MS EXCEL file, and then saved them in your XLS data file. If there are problems with any variables brought in, such as using disallowed characters in the variable name, an error message will be printed in the SPSS output.

Variable Formatting

If you entered your data using MS EXCEL and saved it as an XLS file, then when you open this with SPSS, any formatting you had in the spreadsheet will be carried into SPSS, along with the variable name. When your data is successfully in SPSS there are various formats and labels that can be applied to your variables, or changed from the original values. This is most simply achieved using

the VARIABLE VIEW window shown in Figure 1-4.2. The Variable View window looks very much like the spreadsheet layout except that the columns have headers. The cell entries for each of these columns are:

Name: 8 character name of the variable (note: SPSS 12 allows longer variable names)

Type: Numeric, string, date or several other formats are allowed for variables.

Width: This specifies the maximum number of characters for the column containing the variable value.

Decimals: This specifies the number of decimals for a numeric variable

Label: This Is a descriptive label for each variable. It appears on the SPSS printouts and makes them easier to read than simply the cryptic 8 character variable names.

Values: These are labels that can be attached to specific values of each variable. They are best used on categorical data, for instance sex. Your data may have been coded 1 for male and 2 for female. In order to make analysis outputs more easily read, you may assign value labels such as Male and Female to the variable. When you select the values cell for a specific variable the Value Labels dialog box will come up (Figure 1-4.3). In the example the male label has already been entered and is in the list, the female label is in the process of being entered. When the add button is pressed, female will move into the list. You may enter as many value labels as there are different values for the variable.

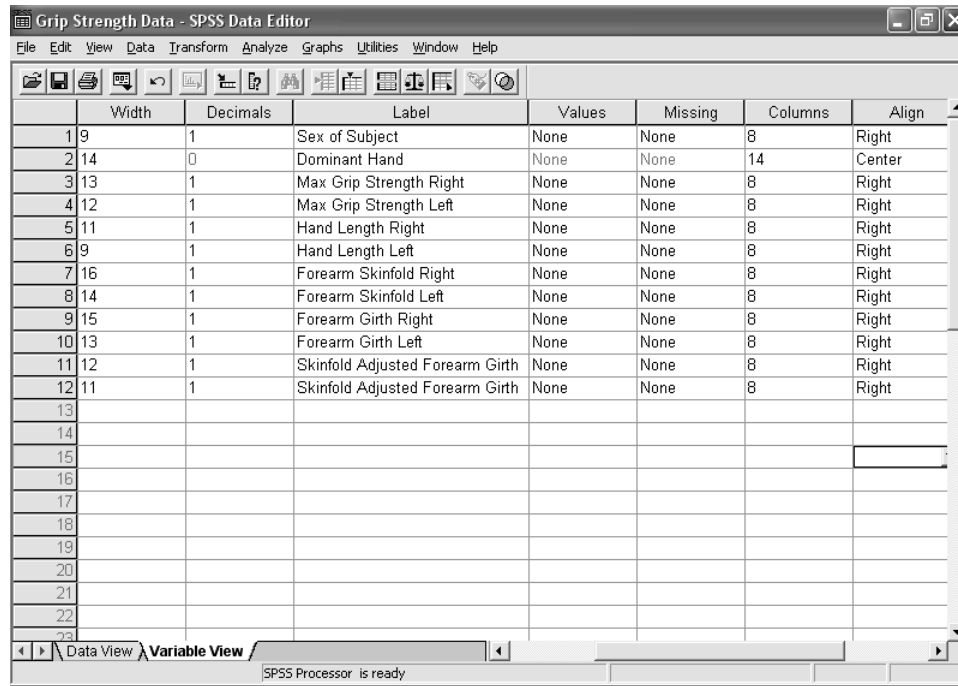**Figure 1-4.1: SPSS Data Editor Window – Data View**

Grip Strength Data - SPSS Data Editor

File   Edit   View   Data   Transform   Analyze   Graphs   Utilities   Window   Help

	Width	Decimals	Label	Values	Missing	Columns	Align
1	9	1	Sex of Subject	None	None	8	Right
2	14	0	Dominant Hand	None	None	14	Center
3	13	1	Max Grip Strength Right	None	None	8	Right
4	12	1	Max Grip Strength Left	None	None	8	Right
5	11	1	Hand Length Right	None	None	8	Right
6	9	1	Hand Length Left	None	None	8	Right
7	16	1	Forearm Skinfold Right	None	None	8	Right
8	14	1	Forearm Skinfold Left	None	None	8	Right
9	15	1	Forearm Girth Right	None	None	8	Right
10	13	1	Forearm Girth Left	None	None	8	Right
11	12	1	Skinfold Adjusted Forearm Girth	None	None	8	Right
12	11	1	Skinfold Adjusted Forearm Girth	None	None	8	Right
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

Data View \ Variable View

SPSS Processor is ready

Figure 1-4.2: SPSS Data Editor Window – Variable View

Value Labels

Value Labels
Value: 2
Value Label: Female
Add   1.0 = "Male"
Change
Remove

OK
Cancel
Help

Variable Type

● Numeric
○ Comma
○ Dot   Width: 12
○ Scientific notation   Decimal Places: 2
○ Date
○ Dollar
○ Custom currency
○ String

OK
Cancel
Help

Missing Values

○ No missing values
● Discrete missing values
   999   998   997
○ Range plus one optional discrete missing value
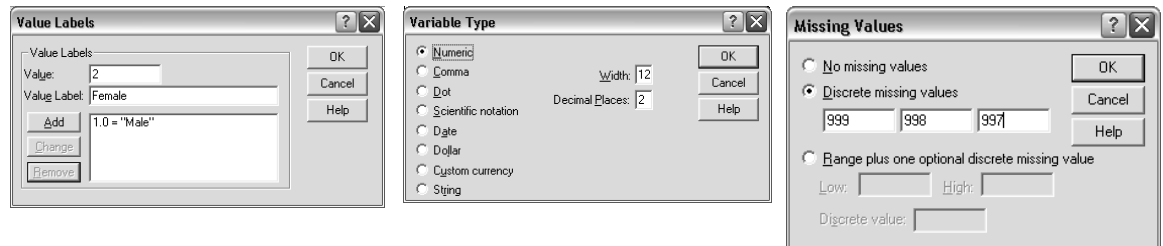   Low: High:
   Discrete value:

OK
Cancel
Help

Figure 1-4.3: SPSS Value labels, Variable Type and Missing Values dialog boxes

Missing: One of the most important formatting steps is to declare what values are to be assigned as missing values. Often you have incomplete data for some of your subjects. It is usual practice to assign numeric codes to missing values and then to declare them using the missing values option. For example, if you were measuring many anthropometric and performance variables on a group of individuals you might be unable to measure body weight in some of the subjects. You would therefore need to assign missing values to body weight for those subjects. This allows for later categorization according to the reason for the missing data. Code the appropriate data field as blank, or a popular convention is to use 9s as missing values codes. For body weight the range of possible values in kgs might be 0 to 150, therefore 999 could be used as a missing value for weight. Sometimes missing values occur for different reasons. If the subject was not present for testing you might code as 999, if the subject refused to be weighed you might code as 998 and use 997 as a code for missing because the subject was in a wheelchair. An important feature of missing values is

that if a compute statement refers to variables with missing values, then the result will be set to a system missing value. When SPSS analysis is run, the print out can include missing values reported for each code. Just remember to use a number of 9s bigger than the highest possible value for the variable. For instance, absolute VO$_2$max, you could use a single 9 as a missing value code, since no one can have a VO$_2$max of 9.0 l.min^{-1}. Zeros tend not to make good missing vale codes as 0 can often mean a real value. SPSS will allow you to code a range of values as missing. For instance you could code 900 to 1000 as missing and any values in that range would be treated as missing values. However, whatever missing values you choose, you must code your variables using the Missing values dialog box.

Columns: This is the maximum character width for the display of the variable value.

Align: This is the alignment of the variable in the column (right, centre, left).

Measure: This denotes whether the variable is scale, ordinal, or nominal in nature.

Data – Select Cases

A very useful command, is the SELECT CASES option on the DATA menu, which allows you to analyze only part of the data set. The dialog box shown as Figure 1-4.4 (left) would appear when you select the DATA menu and click on SELECT CASES. This dialog box allows you to set up the conditions necessary to be satisfied for inclusion in analysis. Note that there is an option to select a random sample from the data set. If the IF CONDITION IS SATISFIED selection is made, and the IF button is clicked, a second dialog box (Figure 1-4.4 right) will be displayed. This will allow the condition for selection to be constructed. In the example, it was select if the variable "domhand" had the value "R", which means use those individuals who declared their right hand as their dominant hand. The condition could be a more complex expression if desired, requiring the solution to an equation that you typed in to this entry box.
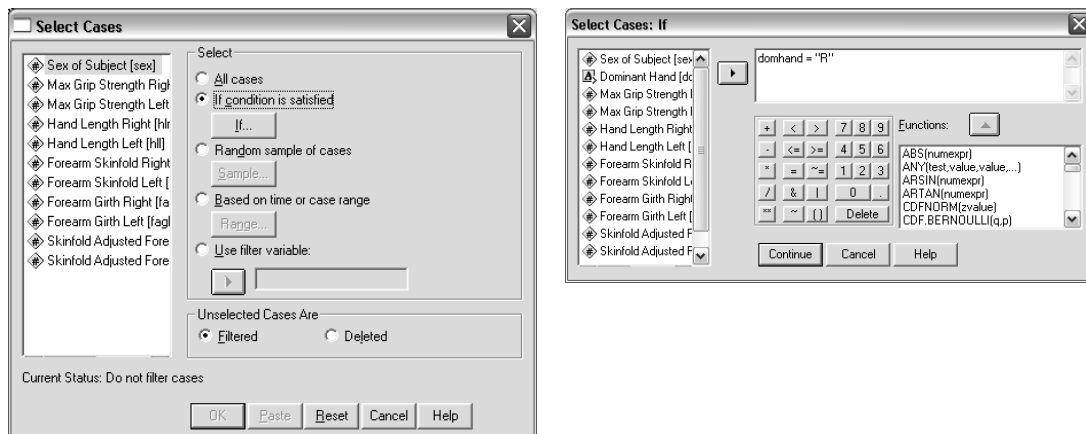

Figure 1-4.4: SPSS SELECT CASES dialog box (left), SELECT CASES: IF dialog box (right)

Data - Split File

When running statistical analyses on a data set, we often want to run the same analysis on different sectors of the data. For instance, we often want to keep males and females separate for analysis. This could be achieved by using SELECT CASES to select males, run the analysis, then select female cases and run it again. This is possible, but very inefficient. There is a more

Figure 1-4.5: SPSS SPLIT FILE dialog box

effective way of achieving this result. In the DATA menu you will find the SPLIT FILE command. Select SPLIT FILE to bring up the dialog box show in Figure 1-4.5. What SPLIT FiLE does is produce a filter variable based upon the variable(s) select to be used for file splitting. There are two options for how the splitting will occur. If COMPARE GROUPS is checked then the results of the statistical procedures for the two or more groups will appear in the same table in the output page. If ORGANIZE OUTPUT BY GROUPS is checked, then the results of statistical procedures will be separated by the splitting variable(s) in the output. The choice of which to use, is simply a matter of preference of the user, the information presented will be the same, just organized differently.

The GROUPS BASED ON entry box can receive more than one variable. In Figure 1-4.5 the variable Sex is listed first then the variable DomHand. What will happen is that the file will be split hierarchically based upon the first and then the second variable. If the variables were listed in the reverse order then the data would be split by DomHand first then Sex.
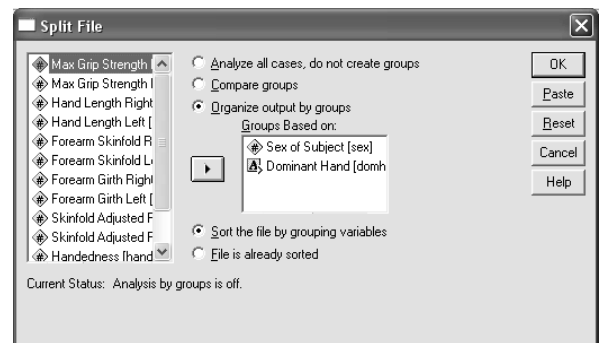
Transform - Compute

Very often you need to create a new variable from a relationship of existing variables. This is achieved using the COMPUTE option of the TRANSFORM menu, which would bring up the dialog box shown in Figure 1-4.6 (left). In this example, a new variable GripDiff is being created using the equation GripDiff = GripR – GripL. The name of the new variable GripDiff is typed into the TARGET VARIABLE box and the rest of the equation is typed into the NUMERIC EXPRESSION box. When the OK button is clicked the new variable will be created and calculated with the new values appearing in the DATA EDITOR window as a new variable on the far right hand side of the data set.
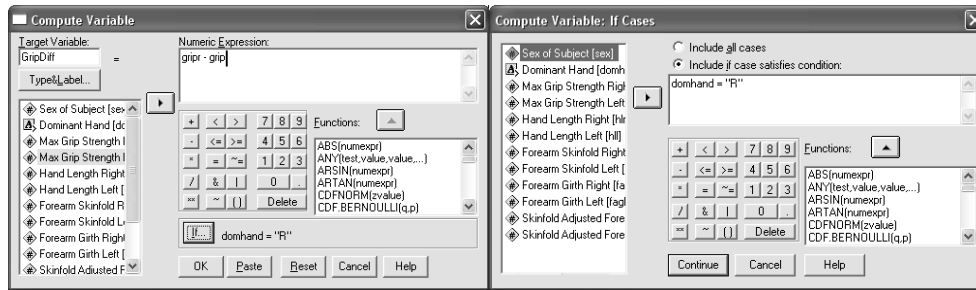
Figure 1-4.6: COMPUTE VARIABLE Dialog Box (left) and COMPUTE VARIABLE: IF CASES dialog box (right)

On the COMPUTE VARIABLE dialog box there is an IF button. If you click this, the COMPUTE VARIABLE: IF CASES dialog box will open. Here you can create variables conditional upon satisfying the IF criteria. In the example, when DomHand equals "R" (note the use of quotes around R to indicate nonnumeric values) then the calculation is GripR – GripL, the other condition entered would be if DomHand equals "L" then the equation would be GripL – GripR. What this will achieve is that the variable GripDiff wil be the grip strength in the dominant hand minus the grip strength in the non-dominant hand, rather than just right hand minus left.

Statistical Tests

Unlike Excel, SPSS was developed specifically for statistical analysis and therefore is a much more sophisticated package. We will be using many statistical analyses in this text, and as each one is used the exact details of SPSS operations will be described. The purpose here is to introduce you to the way in which the statistical analyses can be run. There are dialog box formats that are used for all of the tests. Figure 1-4.7 shows the result of clicking the ANALYZE menu in the DATA EDITOR window. The drop down menu lists all of the general categories of statistical analysis. The example shows the
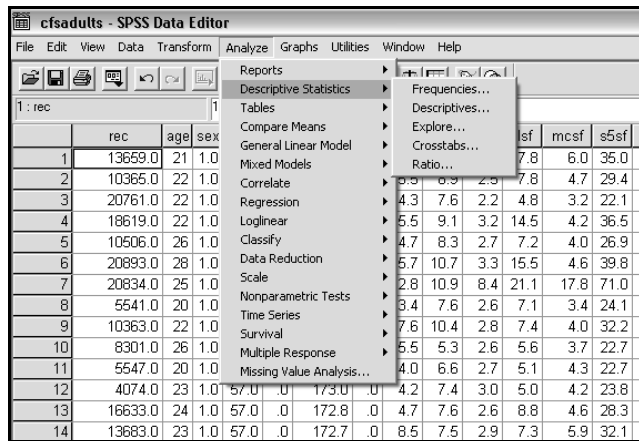

Figure 1-4.7: ANALYZE – DESCRIPTIVE STATISTICS drop down menus

DESCRIPTIVE STATISTICS option being selected, resulting in a drop down menu of five choices of analyses under DESCRIPTIVE STATISTICS. When the DESCRIPTIVES option was selected, the dialog box in Figure 1-4.8 was presented. When a statistical analysis is selected from the drop down menu the dialog box specific to the selected test will be displayed. These tend to operate similarly

Figure 1-4.8: COMPUTE VARIABLE Dialog Box (left) and COMPUTE VARIABLE: IF CASES dialog box (right)

for the different tests. There will be a list of variables in the data set presented on the left with one or more boxes on the right to receive variables designated for testing. When a variable is selected in the left hand list the arrow will blacken and point to the right. If the arrow is clicked the variable selected will be moved into the variable list on the right. In the example, the variable Age has been put into the list.
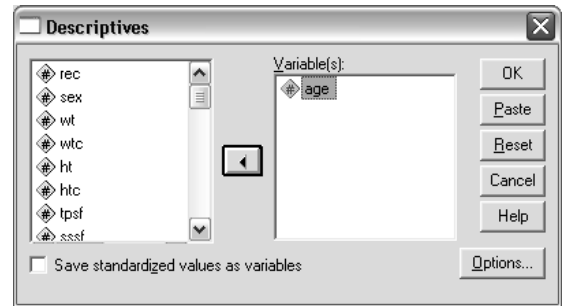
Almost all of the statistical analysis dialog boxes have on OPTIONS button. This will bring up a dialog box that will allow selections about test-specific configurations. In the case of our DESCRIPTIVES example the OPTIONS dialog box is shown in Figure 1-4.9. Here you can select

Figure 1-4.9: DESCRIPTIVES: OPTIONS Dialog Box

which statistics will be calculated and the order of presentation of results. What is contained in the OPTIONS dialog box is specific to the test being used.
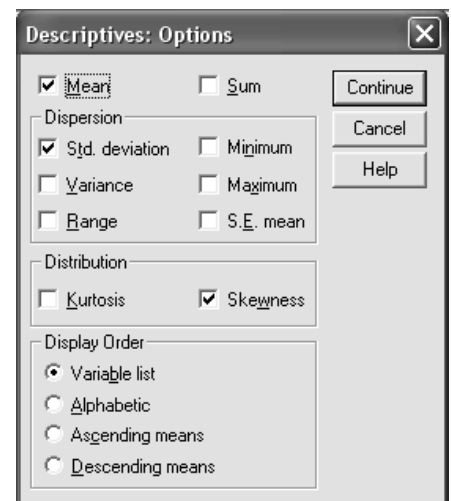

Descriptives

Descriptive Statistics

	N	Mean	Std.	Skewness	
	Statistic	Statistic	Statistic	Statistic	Std. Error
AGE	10909	38.334	13.290	.551	.023
Valid N (listwise)	10909				

Figure 1-4.10: DESCRIPTIVES output of Descriptive statistics for Age in the whole Canada Fitness Survey adults data set (N = 10909)

When the test is run the results are sent to the OUTPUT window. Figure 1-4.10 shows the output for the DESCRIPTIVES dialog box shown in Figure 1-4.8. It simply gives the selected descriptive

statistics of Age for all subjects (N = 10,909) in the CFS (Canada Fitness Survey) adult data set. The output can be printed or component parts can be copied and pasted into WORD files as desired.