Chapter 0. Introduction to SPSS, an IBM Company[†]

Topics covered in this chapter:

- **Introduction**
- **Accessing SPSS**
- **Opening and Saving Data Files**
- **Defining Variables and Entering Data**
- **Opening Excel Files**
- **Recoding Variables**
- **Deleting/Inserting a Case or a Column**
- **Selecting Cases**
- **Using SPSS Help**

Introduction

In this chapter we introduce SPSS, the Statistical Package for the Social Sciences. This manual is intended to help a student perform the statistical procedures presented in your W. H. Freeman text: The Basic Practice of Statistics. This supplement is intended for use with SPSS for Windows version 16.0 (current at this writing). However, the instructions included here will work for most versions of SPSS and for most basic statistical procedures.

Throughout this manual, the following convention is used: commands you click are in boldface (e.g., go to **File**), variables are in boldface and italicized (e.g., ***Count***), and text you type is in quotes (eg. "probability").

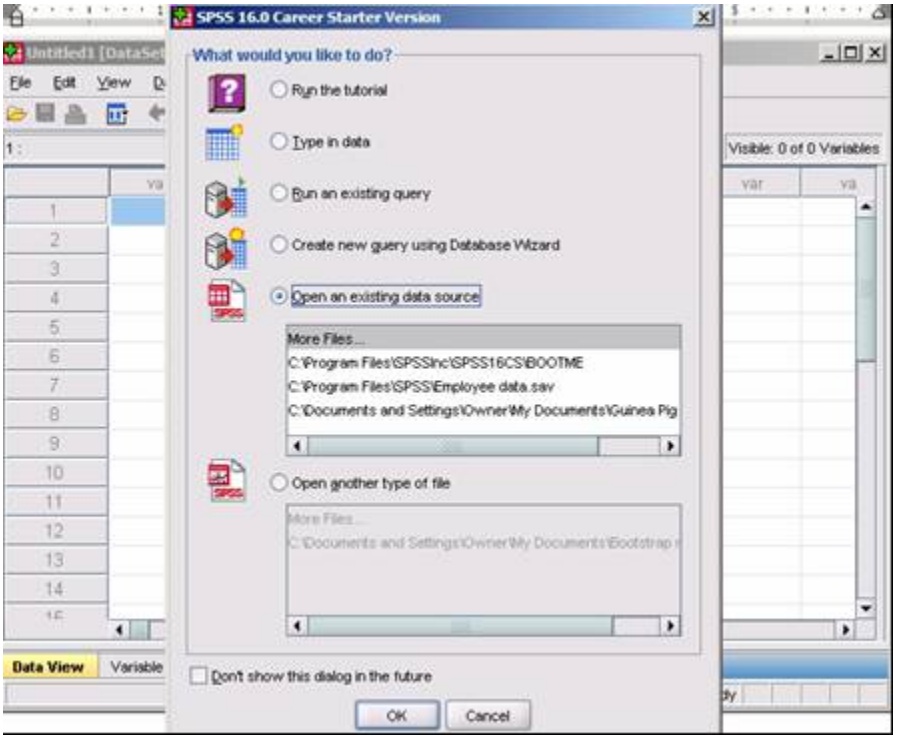
Accessing SPSS

If you work in a lab, locate SPSS in the computer. You should look for the following icon on the desktop:

SPSS 16.0.lnk

[†] SPSS was acquired by IBM in October 2009.

Your computer may have a similar icon with an earlier version number. Double click on the icon to start SPSS running. If there is no desktop icon, use the **Start** menu on your computer and open **All Programs**, locate the **SPSS Inc** folder, and follow it to find the program. Once the program has been started, you will briefly see an introductory screen (similar to an Excel or Word start-up screen) followed (most likely, unless this has been disabled by checking the box at the bottom) by this screen.

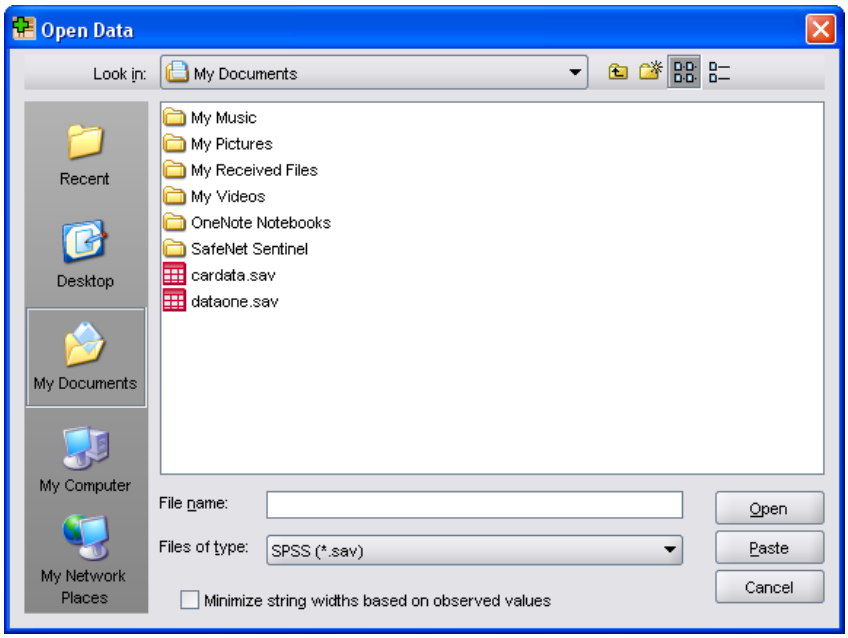


Two SPSS windows will actually be open at this point — the output viewer and the Data Editor. Before doing any statistics or graphs, we must have data. The purpose of the introductory screen is for the program to determine your data source. Behind the data source selection box is a blank spreadsheet — the **Data View** screen.

Since all data sets used in your text are on the included CD-Rom (and on the Companion Website), here you will most likely click **OK** (or press **Enter**) to select the default option which is to **Open an existing data source**.

Opening and Saving Data Files

If the introductory data source selection menu is not presented (or you want to proceed to another data set within an SPSS session), click **File** , **Open** , **Data**. Initially, you will see the screen below.



File selection works much the same as any other Windows program. In the **Look in** box, select the location of your data set (drive and folder). The SPSS default data file extension is .sav. Data files on the CD-Rom are saved as SPSS portable worksheets, so change the box labeled **Files of type** to SPSS Portable (.por) as in the screen following.

File naming conventions used on the CD and Website are the following:

1. The first two characters describe the type of data set. Examples are eg, exercises are ex, figures are fg, and tables are ta.
2. The second two characters indicate the chapter number.
3. Numbers after the dash correspond to the example, exercise, figure, or table number within the chapter.

Once you have located the file you want, click **Open**.

To save an SPSS data file, click **File** then **Save As,** or click **File** then **Save All Data** depending on your SPSS version. In the File Name box, type the name you wish to give your data. The default folder is PC-SPSS. Be sure to change that if you want a different location, such as diskette or flash drive.

Defining Variables and Entering Data

In the event you need to enter your own data for a project, on the first (opening) screen select the **Type in data** button and click **OK**. You will see the blank SPSS **Data Editor** spreadsheet window.

Example 0.1: Creating an SPSS Data File by Entering Data.

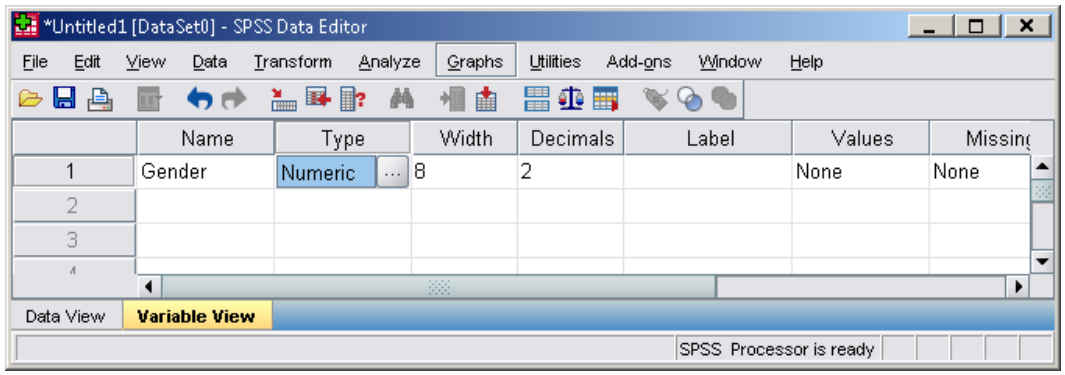
The following data set contains 10 randomly selected scores in the final exam of a basic statistics course at XYZ College. Along with the final exam scores, the number of classes missed during the semester and the gender of the students were also recorded. The data set is given below:

Gender	Number Classes Missed	Final Score
Male	2	83
Female	0	93
Male	6	61
Female	1	73
Female	0	95
Female	4	75
Male	3	77
Male	4	71
Female	5	68
Female	4	59

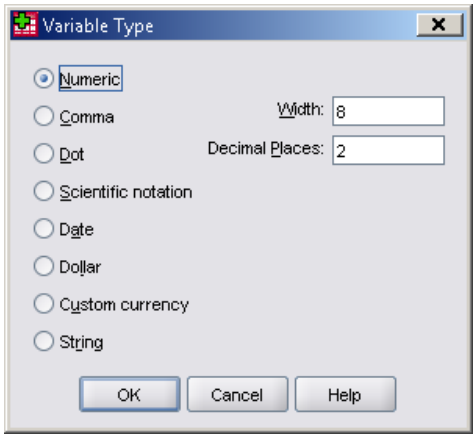
You can simply begin either by typing the data into the spreadsheet, or by defining the variables. For completeness sake, both steps should be completed, but order is unimportant.

To define the variables, click on the **Variable View** tab at the bottom of the **Data Editor** window.

Under **Name**, type the name of the first variable (eight characters or fewer, beginning with a letter or the underscore sign). In this case, the name of the first variable is ***Gender***. Press the **Tab** key to advance to the **Type** box. Notice what SPSS defaults variables to: Numeric, with two decimal places, occupying eight columns. Gender is a categorical variable. We need to change this.



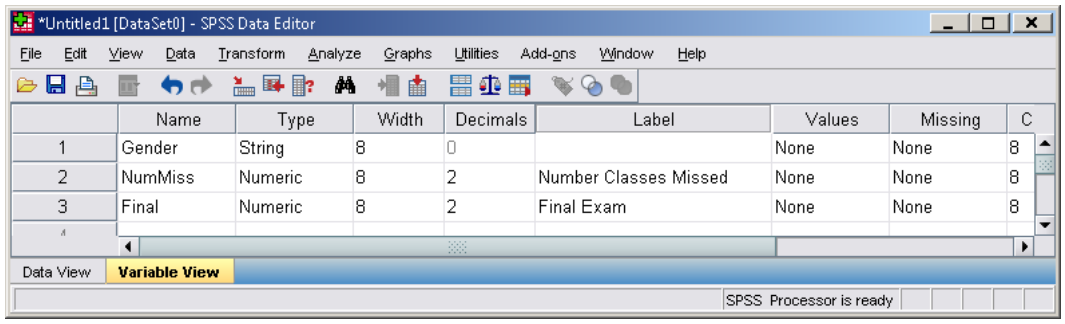
To change the variable type, click in the highlighted box, then click on the small button that appears at the right.



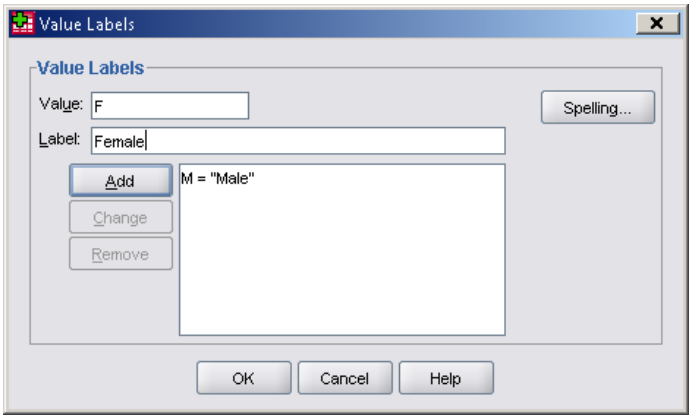
To change this variable to a categorical one, click on **String**. You will be allowed to change the maximum number of characters (the default is eight) if desired. When finished, click **OK**.

Enter the name of the second variable ***NumMiss***. This variable name could stand for some explanation on output, rather than just this cryptic name. Press the Tab key to move to the **Label** column. Type in a more appropriate "long" variable name, such as "Number Classes Missed".

Enter the third variable name ***Final*** and label it as "Final Exam Score" as just detailed. At this point, our variables definition should look like that below.

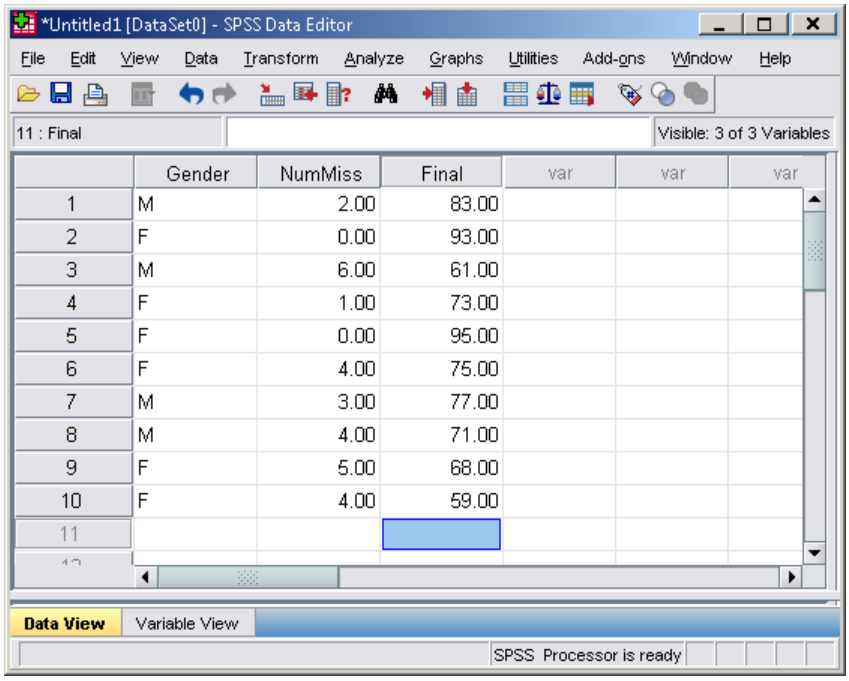


Lastly, consider the ***Gender*** variable. We'd like our data entry to be as easy as possible, but have SPSS print out the full word for Male and Female students. If we just want to enter "M" (or "F") we can define value labels that will print the full descriptor. Click the cursor in the **Values** field of the ***Gender*** variable. A small box like will appear as before. Click it to get a dialog box.



Here, I have already added the label for "Male" and input both the value and label for "Female". To add the value label, click **Add**. When finished adding labels, click **OK**.

To start entering the data, click on **Data View** and enter the values, pressing **Tab** after each entry. The program will automatically advance to the next row after the third variable for an individual has been entered. If you make a typographical error, simply click on the cell and type in the correct value.

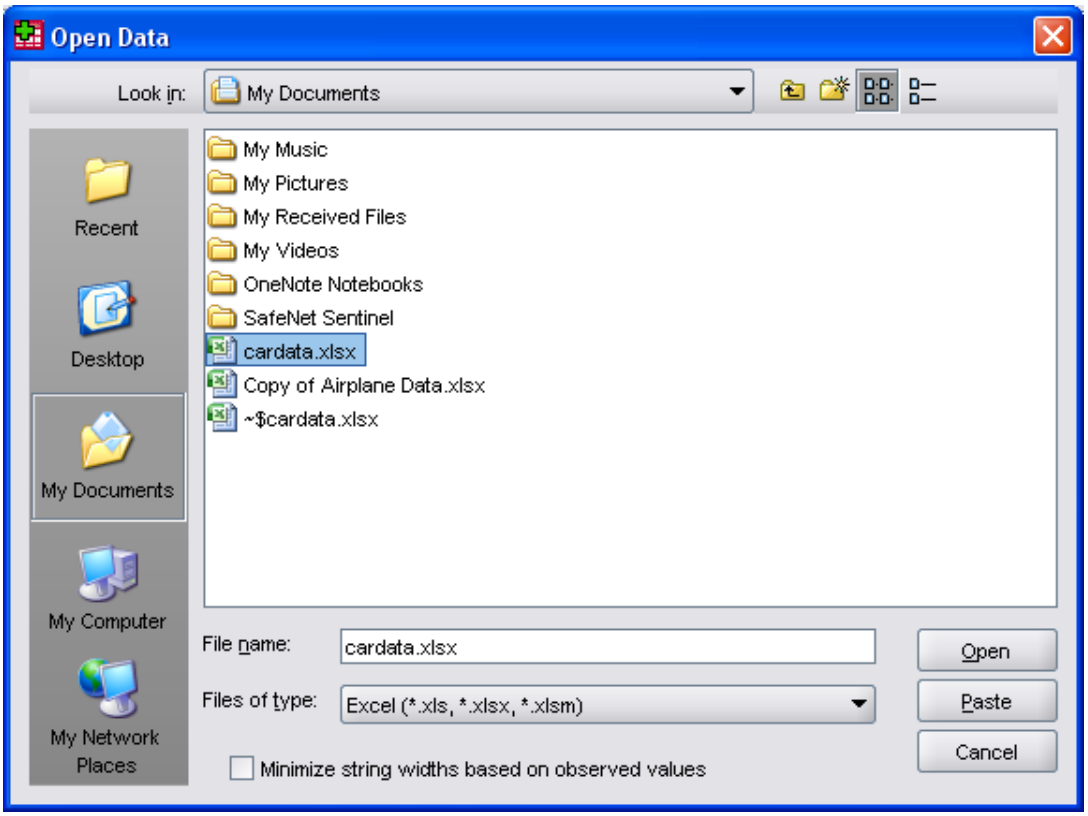


To save your data in an SPSS formatted file, follow the instructions in the preceding section.

Opening Excel Files

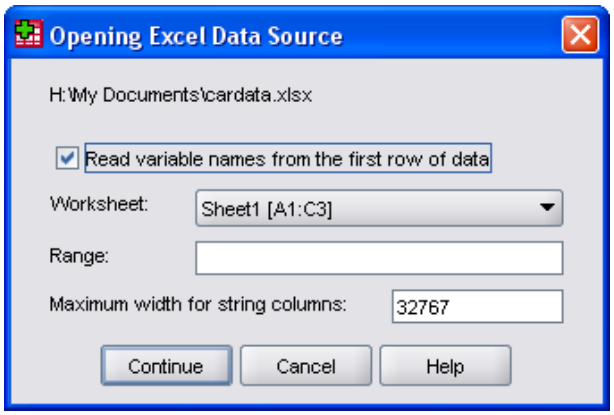
To open an Excel data file, follow these steps:

1. Click on **File**, **Open**, **Data,** then the **Open Data** window will appear as already shown.
2. Choose the directory or location where the desired file is located. In our case, the file is stored in the My Documents folder.
3. Change the **Files of type** box from the default .sav to Excel (.xls) or all files.



4. Click on the desired file name. In our case, it is "cardata.xls" as shown above.

Depending on the Excel file, you need to know whether the names of the variables are located in the first row or not, i.e., where do the actual data start? Is it in the first or second row? In our case, the first row does have names, so leave the **Read variable names from the first row of data** box checked. Click **Continue** to open the file.



SPSS will take most variable attributes from the information in the Excel file. You probably will want to give more meaningful "long" variable names. Click on **Variable View** and add **Labels** as shown above.

Recoding Variables

One can change a categorical (string) variable to numeric. Also, one can transform a quantitative variable from one form to another by categorizing or by recoding the variable. The following example shows how to categorize a numeric variable in SPSS.

Example 0.2: Recoding Variables.

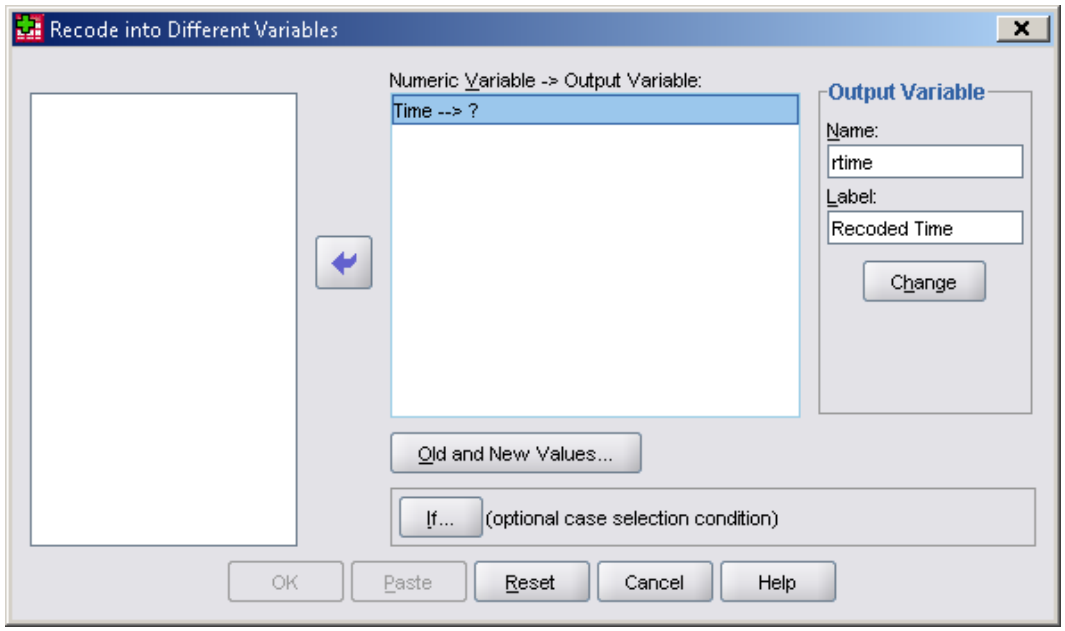
The following data represent the waiting time (in seconds) for a random sample of 30 customers at a local bank.

49, 160, 80, 220, 170, 92, 178, 66, 124, 144, 71, 183, 248, 191, 155, 166, 256, 300, 180, 166, 171, 280, 144, 110, 267, 188, 160, 90, 205, 136

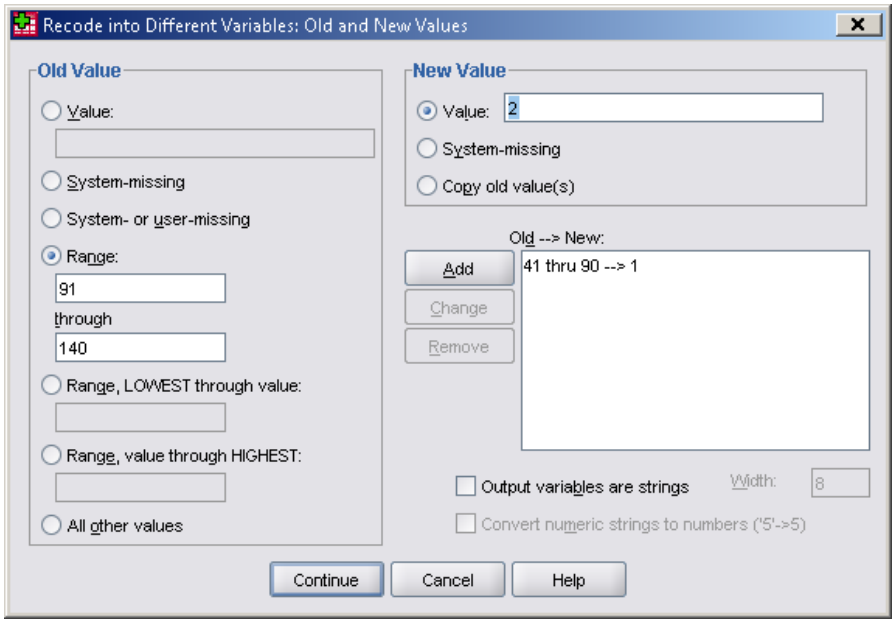
It may be more useful to group these data into non-overlapping classes (i.e., to create frequency tables). Let us recode these data to six equal width classes. The number of classes is usually determined by sample size and should fall between 5 and 20 intervals. A good rule of thumb is to use the square root of the sample size as a rough estimate for the number of classes.

The names of these classes will be 1, 2, 3, 4, 5, and 6. All observations between 41 and 90 seconds (inclusive) will be assigned to class 1, all observations between 91 and 140 will be assigned to class 2, all observations between 141 and 190 to class 3, and so on. Here is how to do it in SPSS.

Click on **Transform**, **Recode into Different Variables**. Time has already been highlighted as the input variable (since it's the only one in this spreadsheet). If there are more variables in your sheet, click to select the one of interest. Click the arrow to move Time (or your selected variable) to the working area at right.



Name the new output variable in the box at right, and give it a "long" name or label if desired. Click the **Change** button to record the new variable name. Click on the **Old and New Values** button. I have already defined the first category displayed in the box at right. Here, I am defining the second category as including the range 91-140 with new value 2. Click the **Add** button to complete this category definition, and define the others.



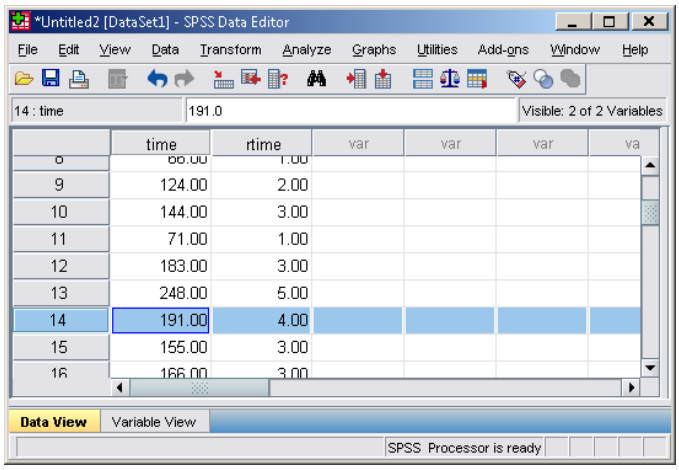
When all categories have been defined, click **Continue** to return to the first **Recode** box. Click **OK** at the bottom of the box to create the new variable.

It will be useful to tell anyone who looks at your output what these recoded values represent. To do this, click on the **Variable View** tab at the bottom of the worksheet, then click in the **Values** box and add value labels as discussed previously.

Deleting/Inserting a Case or a Column

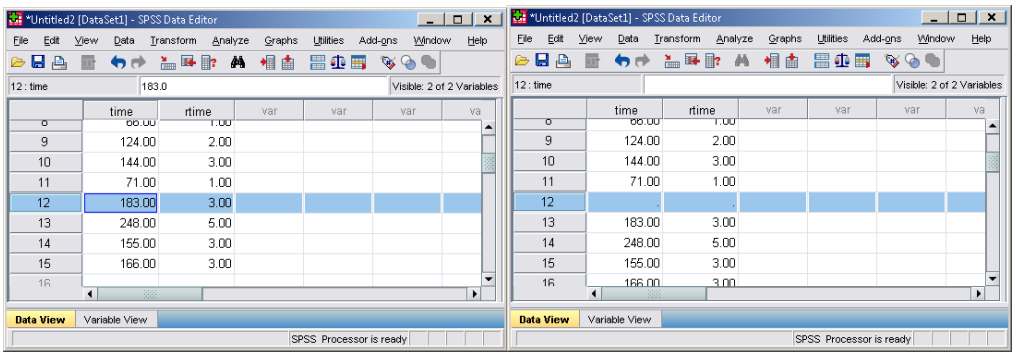
The data presented in the previous example will be used to illustrate the points of this section. To delete a case (an entire row of data), follow these steps:

1. Locate the case to be deleted by scrolling through the data.
2. Click on the case number at the left. The entire row will be highlighted. Suppose observation 14 should be deleted because it was an extra.
3. Press **Delete** on the keyboard.



To insert a case, follow these steps:

1. In the **Data View** window, click on the case number below where the new case should be.
2. Click **Edit**, **Insert Cases**. A blank row will be inserted.
3. Type in the desired data values for that observation.



To insert a new variable it is easiest to define one in the **Variable View** as previously described after those already in the data set. If you want to insert one in the middle, follow these steps:

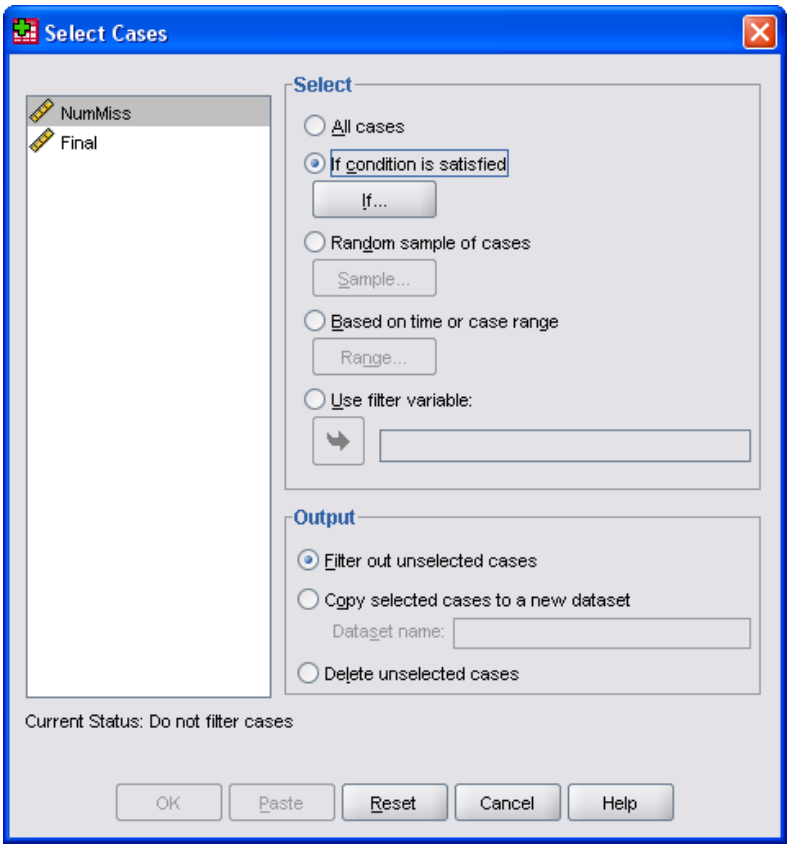
1. In the **Variable View**, click on the variable below where the new one is to be inserted.
2. Click on **Edit**, **Insert Variables**. A new variable with name of the form ***VAR00xx*** will be inserted. Change this name to the desired name, and also change any of the default characteristics (variable type, number of decimal places, etc.) as needed.

To delete a variable within the **Data View**, click on the variable name and press the **Delete** key on the keyboard.

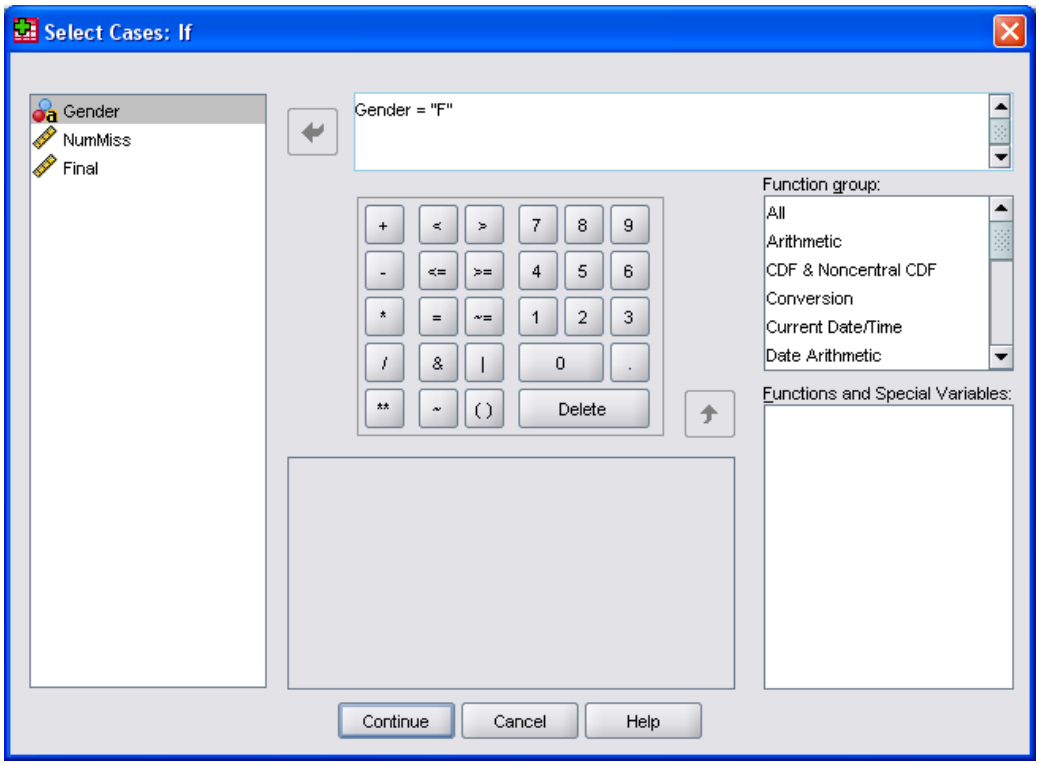
Selecting Cases

Statistical analyses are sometimes needed for part of the data rather than for the entire data set. For example, it may be desired to compare the "Females" against the "Males" for the data on absences and final exam scores used in Example 0.1 (page 4). We might also want to do a regression with and without outliers to examine their impact.

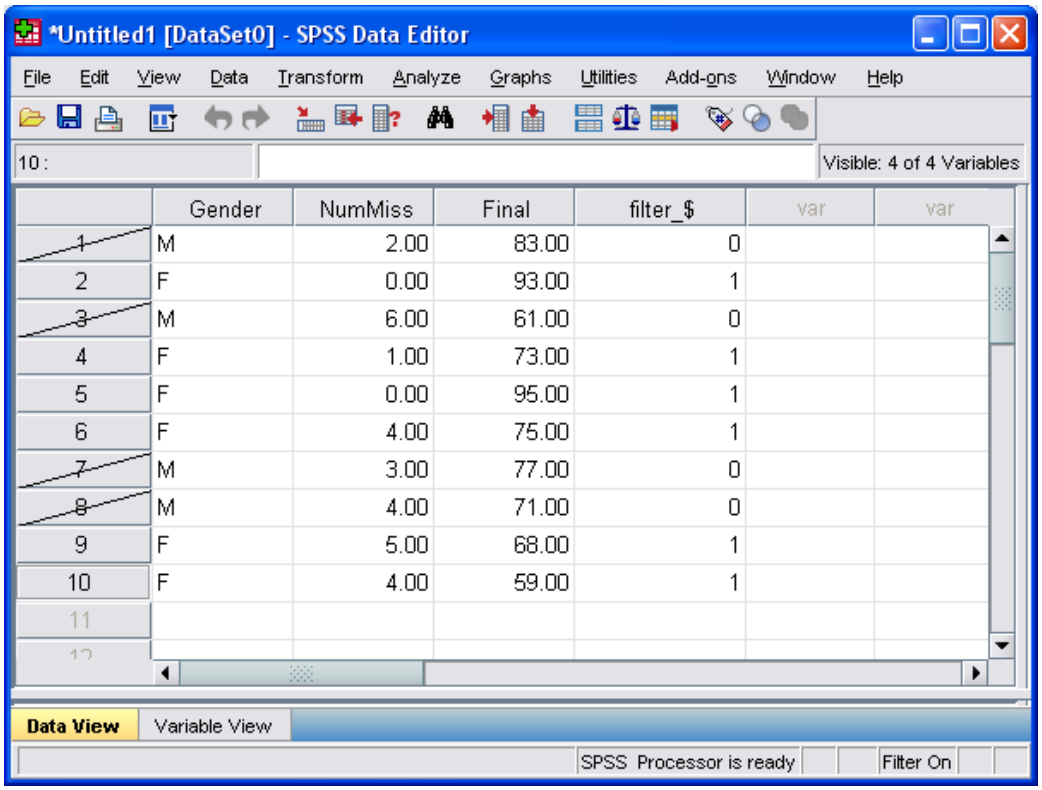
1. Click **Data**, **Select Cases**.
2. Select the option for **If condition is satisfied** and click the **If** button.



3. Highlight the variable name to be used and press the right arrow box to transfer this into the condition box. Complete the condition (in this case we want to select "Females"). Click **Continue** to return to the main **Select Cases** box.



4. Click **OK** to perform the selection. We see in the screen at right that "Males" will now be ignored, and a new variable named ***filter_$*** has been created. This variable has values 0 or 1 according to whether the case has been excluded or not.

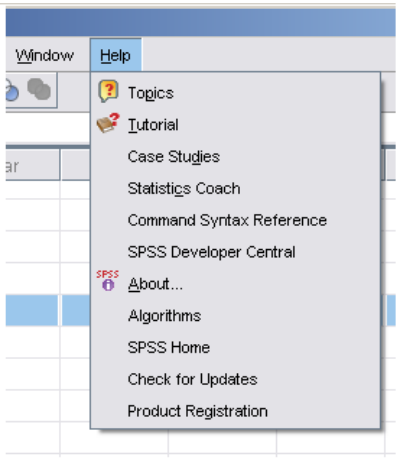


	Gender	NumMiss	Final	filter_$	var	var
1	M	2.00	83.00	0		
2	F	0.00	93.00	1		
3	M	6.00	61.00	0		
4	F	1.00	73.00	1		
5	F	0.00	95.00	1		
6	F	4.00	75.00	1		
7	M	3.00	77.00	0		
8	M	4.00	71.00	0		
9	F	5.00	68.00	1		
10	F	4.00	59.00	1		
11						

To return to using all cases and remove the filter, go back to **Data**, **Select Cases** and select the **All Cases** button, then click **OK**.

Using SPSS Help

Suppose you were looking for information on how to do something in SPSS and you can't find it in this manual. Help is available in several forms by clicking **Help** on the right-hand side of the top menu bar. Context specific help is available in every dialog box simply by clicking the button.



The **Tutorial** offers basic information on certain topics, in much the same manner as this manual. The **Statistics Coach** presents a series of screens to narrow down the search of topic and presents sample output as well. Lastly, one can search for help by **Topics**. This author recommends that if searching by topic, select **Index** after the initial **Topics** selection. Enter the topic name in the search box. As you type more characters, the index at left will move to try to "zero in" on the topic of interest. When you see it, highlight the topic name and click **Display**. The screen below illustrates the initial results from a search for t-tests.