1. Introduction to SPSS for Windows

SPSS (Version 16) is a statistical software package available in many of the campus computer labs on a Windows platform for PCs. This manual serves to explain how to use SPSS to perform some basic types of statistical analyses. This manual occasionally refers to "the text". This means the textbook for Math 241 which is *Workshop Statistics: Discovery with Data, 3rd Edition* by Allan Rossman and Beth Chance.

1.1 Starting SPSS

Figure 1-1 shows the opening screen in SPSS 16.0. In the section labeled "Open an existing file", the most recent files used in SPSS are listed and they can be opened by clicking on the desired file name and clicking **OK**. You can scroll down the list of recent files by clicking ▼. If you want to open an existing SPSS file not in the list (for example, one on a floppy disk) then click **More Files** and then click **OK**. If you want to type in data by hand into a new file, click on "Type in data" and then click **OK.** We will not use any of the other options in this window.

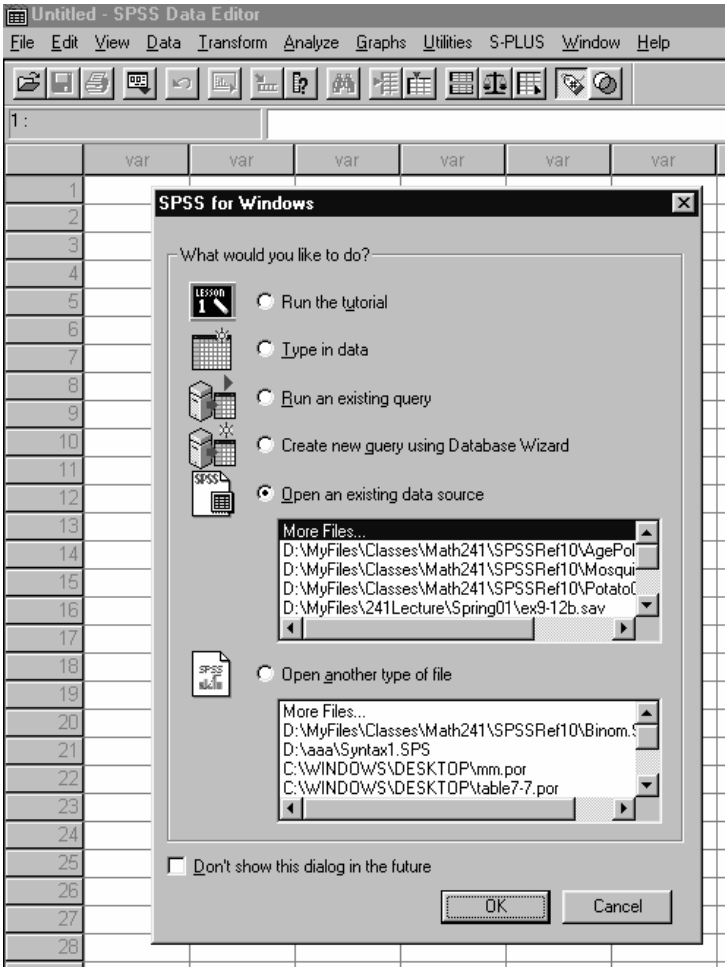


Figure 1-1

The window in the background of Figure 1-1 is the SPSS Data Editor. At the bottom of the Data Editor are two tabs labeled Data View and Variable View: 

. Click on the desired tab to switch to that view. When Data View is selected (the default), the Data Editor shows a spreadsheet where columns represent variables and rows represent individual cases or observations. Here, you may enter data, edit existing data, or delete data. When Variable View is selected, the Data Editor shows a spreadsheet which contains information about the variables (see the example in Figure 1-2 on page 6). In this view, each row represents a variable and each column a property of variables. The meanings of the different column names are given in Table 1-1. To change a characteristic, click on the appropriate cell. Except for the "Name" and "Label" columns, one of these two buttons will appear in the cell:  or  . If it's  , click on it and a screen will appear which allows you to change the characteristic. If it's  , then clicking on the arrows will change the number in the cell (for instance, to change the number of decimal places displayed). The characteristics of the variables can be changed at any time (before or after the data have been entered). Variable labels ("Label") and value labels ("Values") are discussed in separate sections later in this chapter.

Meanings of columns in Variable View

Column	Meaning
Name	Variable name (no embedded spaces allowed)
Type	Numeric is default; if values are names (e.g., state names) then select String
Width	Maximum number of places for each data value
Decimals	Number of decimal places displayed for numeric values
Label	Variable label: you can enter a longer, more descriptive variable name; see example below
Values	Value labels: can attach a name to each possible value of a variable (e.g., 0=Female, 1=Male); see example below
Missing	If missing values are coded by a specific value (e.g., -999) you can indicate that here
Columns	Width of column displayed in Data View (does not affect the data values)
Align	Alignment of values as displayed in Data View: Right, Left or Center
Measure	Variable type: Scale (=quantitative), Nominal (=categorical), or Ordinal

Table 1-1

The SPSS Menu bar (**File, Edit, View,…,Help**) appears near the top of the Data Editor. Each of these main menu items contains a submenu of options.

In addition to the Data Editor, the other primary window is the Output window (actually, it's named "Output1 - SPSS Viewer"). This window contains all of the output generated as you run various statistical analyses and request various types of statistical information. Initially, this window will not be present, but will appear once you perform some statistical analysis requiring output. Every time a statistical request (either numerical or graphical) is made, the resulting output appears in the Output window. To move between the Data Editor and Output windows, select **Window** from the SPSS Main Menu and then click on the name of the desired window. The current active window has a check in front of that window name. Until you have created output, the Output window will not appear on the list of windows.

1.2 Types of Variables

The text (Topic 1) divides variables into two main types: quantitative variables and categorical variables. A **quantitative** variable (also called a **measurement** variable) is one which records the quantity of something and is always numerical. A **categorical** variable records into which of two or more categories each case falls. SPSS has different names for these. It calls a quantitative variable a **scale** variable (quantitative = scale). It calls a categorical variable a **nominal** variable (categorical = nominal). SPSS assumes that a variable is a quantitative variable unless you change the variable type. Changing the variable type is described in Section 1.4 of this manual.

Note: SPSS has one more type of variable: **ordinal**. An ordinal variable is a categorical variable where the order of the categories is important, such as the quality of mail service (good, adequate, poor) or class standing (freshman, sophomore, junior, senior). SPSS treats nominal and ordinal variables the same for virtually all purposes. We won't need the ordinal type for this course so you can safely ignore it and always choose "scale" or "nominal".

1.3 Character (String) Variables

Sometimes, you will want to enter data values that are not numbers but are character strings (a "string" is a sequence of letters, numbers and other characters). For example, if I had a data set that consisted of the population of each U.S. state, I might want to have two variables, the name of the state and its population. The name of the state would be a string variable. SPSS allows you to enter character strings as data values.

For some character variables, it is easier to enter the variable as a quantitative variable than as a string variable. For example, if the sex of a person is a variable, then you might want to enter "Female" and "Male" as the possible values; however, that would require lots of typing. An alternative is to use numerical values, such as 0 and 1, to indicate sex, and then to use **value labels** (Section 1.9) to indicate that 0 represents "Female" and 1 represents "Male". This latter method saves time in entering the data and reduces the possibility of typing errors. It also allows you to specify the order in which the categories of a categorical variable will appear in SPSS output. For example, suppose that each student in a class is asked to identify his or her political views as Liberal, Moderate or Conservative. If this variable is entered directly as a string variable with values "L", "M" and "C", then SPSS will always display the categories in alphabetical order (for example, in tables and bar graphs), which is not the natural order. In order to force SPSS to display the categories in the natural order, one should code the data as 1 (for Conservative), 2 (for Moderate) and 3 (for Liberal). Then use value labels to indicate that 1 represents Conservative, etc. SPSS will then display the categories in numerical order but will replace the numerical codes by the value labels in all output.

String variables are most useful when there is a unique identifier for each case, such as the name of each state. In most other cases, it is better to enter the variable as a quantitative variable and use value labels. See the end of Section 1.4 for instructions on entering string data.

1.4 Entering Data in the Data Editor

Many of the data files in this course will be made available to you at the class web site as SPSS data files that can be downloaded. However, you may occasionally need to enter some data by hand. This you will do in the Data View of the Data Editor.

Example 1-1: Student Measurements. The sex, foot length, height and arm span (all in centimeters) of each student in a class of 25 students were recorded. Sex is a categorical variable

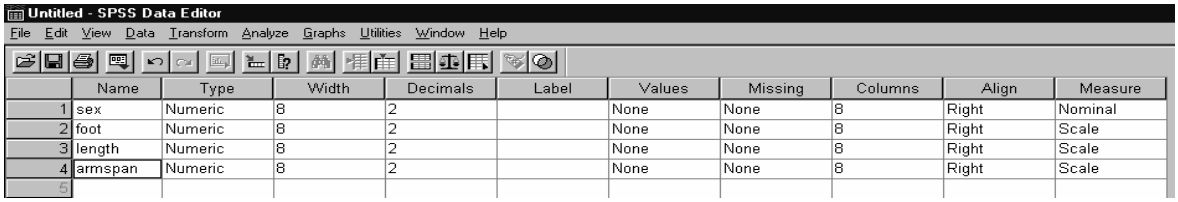
and foot length, height, and armspan are quantitative variables. The first five cases are presented in Table 1-2.

Student	Sex	Foot length	Height	Arm span
1	F	24	162	166
2	F	23	161	160
3	M	25	171	167
4	M	25	166	167
5	M	26	173	168
…				

Table 1-2

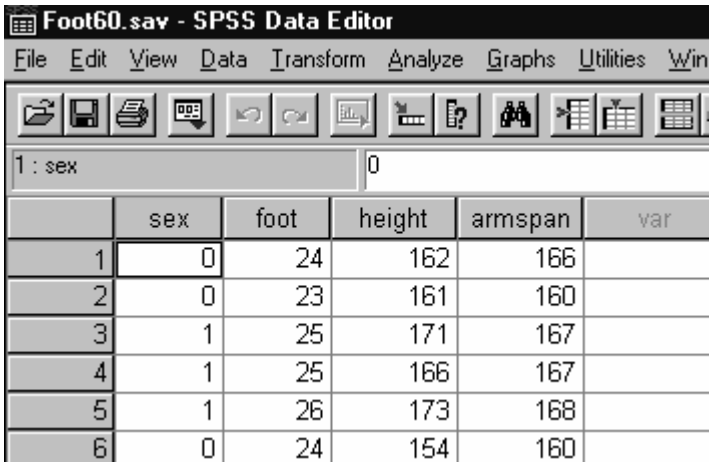
To create this data set in SPSS, follow these steps:

1. Either select **Type in Data** from the "SPSS for Windows" window when you first start SPSS and click "OK", **or** click **File**, click **New** and click **Data** at any time during your SPSS session.
2. There are two tabs at the bottom of the blank spreadsheet that appear, one labeled "Data View" and the other labeled "Variable View". Click on "Variable View".
3. Enter the name of the first variable, **sex**, in the "Name" column in row 1.
4. Enter the names of the remaining variables in the "Name" column for the next three rows. Since variable names may not have any embedded blanks, we'll use the names **foot**, **height**, and **armspan**.
5. The remaining columns of the "Variable View" spreadsheet contain characteristics of the variables that can be changed. Since ***sex*** is a categorical variable, change the value in the "Measure" column (last column) to "Nominal" instead of "Scale". The "Variable View" will now appear as in Figure 1-2.
6. Return to Data View by clicking on the Data View tab and enter the data values. The variables ***foot***, ***height***, and ***armspan*** are all numeric and so the values can be entered directly. The values of the variable ***sex*** are not numeric and can be entered in two ways. The preferred way is to assign numerical values, such as 0 and 1 and then add **value labels** (Section 1.9) to indicate that 0 represents "Female" and 1 represents "Male". That is what we'll do here (we also have to make sure to change ***sex*** to a "nominal" variable as outlined in the previous step). However, you can also enter character data directly. To do this, you must change the "Type" of the variable from "Numeric" to "String" in the "Variable View" window; see below.
7. When you are done, the spreadsheet should look as in Figure 1-3.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	sex	Numeric	8	2		None	None	8	Right	Nominal
2	foot	Numeric	8	2		None	None	8	Right	Scale
3	length	Numeric	8	2		None	None	8	Right	Scale
4	armspan	Numeric	8	2		None	None	8	Right	Scale
5										

Figure 1-2



	sex	foot	height	armspan	var
1	0	24	162	166	
2	0	23	161	160	
3	1	25	171	167	
4	1	25	166	167	
5	1	26	173	168	
6	0	24	154	160	

Figure 1-3

Character (String) Variables

Sometimes character variables are necessary because you may have a unique name for each case in the data set, for example, the names of the states in a data set where the 50 U.S. states are the cases. Character variables are also called "string" variables. You must do the following in order for SPSS to accept string input for a variable: change to the "Variable View" window, then click on the **Type** cell (second column) for the variable. Click on ... The window in Figure 1-4 appears. Click **String**. The value in the **Characters** box (which will appear after you click **String**) is the maximum length of a string. Set this value so that it is at least as large as the length of the longest string in the data set.

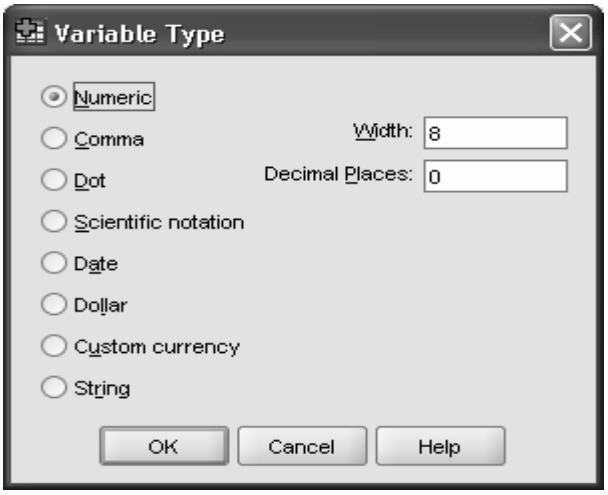


Figure 1-4

1.5 Editing Data in the Data Editor

To change a data value, simply click on the data cell and type in the new value, then press **Enter**. To delete a case (row), select the row by clicking on the row number on the left-hand side of the spreadsheet, then press the **Delete** key on the keyboard (or click **Edit** and **Clear)**. To delete a variable (column), select the column by clicking on the column heading (the variable name), then press the **Delete** key (or click **Edit** and **Clear)**. Variable names and properties can be changed at any time by switching to "Variable View".

1.6 Saving an SPSS Data File

To save an SPSS data file, follow these steps.

1. Click **File** and click **Save as**. The "Save Data As" window in Figure 1-5 appears.
2. If you wish to save the file to a floppy disk, click ▼ by the "Save in" box until **3½ Floppy (A:)** appears, and then click on this name. If you prefer to save the file in a different location, click on ▼ until the name of the desired location appears and then click on this name.
3. In the "File name" box, type in the desired name of the file you are saving. By default, SPSS adds the **.sav** extension to data files.
4. Click **Save**.

Figure 1-5

1.7 Opening an Existing SPSS Data File from Disk

To open an existing SPSS data file from a floppy disk, follow these steps.

1. Either select **Open an Existing File** in the "SPSS for Windows" window when you first start SPSS (see Figure 1-1, page 3) or click **File** and click **Open** at any time during your SPSS session.
2. In either case, all files with the **.sav** extension will be listed for whichever location you choose. Click on the name of the data file you wish to open and then click **Open**.

Some of the class data sets may be saved on the class web site in a special SPSS portable format that has the **.por** extension. You can open a **.por** file just like an **.sav** file, but you must select "SPSS portable (*.por)" in the "Files of Type" box in order to get SPSS to show the **.por** files in the file list.

You can also open **.sav** and **.por** files that are on the desktop by clicking and dragging them onto the SPSS program window.

1.8 Variable Labels

SPSS allows only a maximum of 8 characters for variable names. Sometimes, a longer, more descriptive name is desired. In this case, you can add a **variable label** to a variable name. For example, in **Example 1-1: Student Measurements** (page 5), suppose we would like the variable ***foot*** to have the more complete description "Foot length". Variable labels are entered in the "Variable View" under the "Label" column (see Table 1-1, page 4, and Figure 1-2, page 6). So simply click on the "Variable View" tab at the bottom of the data spreadsheet and enter desired labels in the appropriate cells. Variable labels will appear in all SPSS output.

1.9 Value Labels

We entered ***sex*** as a numerical variable with 0 for Female and 1 for Male in **Example 1-1: Student Measurements.** To have the categories appear as Female and Male in the output, you can add value labels. Value labels are added in "Variable View" under the "Values" column (see Table 1-1, page 4, and Figure 1-2, page 6). The steps for doing this are as follows.

1. Click on the "Variable View" tab at the bottom of the data spreadsheet.
2. Click on the cell in the "Values" column for the variable ***sex***.
3. Click on the button  that appears. The "Value Labels" window will appear as shown in Figure 1-6 except that the box at the bottom will be empty.
4. Type **0** in the "Value" box, then press the **Tab** key. Type **Female** (no quotes) in the "Value Label" box, then click **Add**. The cursor will return to the "Value" box. Type **1**, then press the **Tab** key. Type **Male** in the "Value Label" box, then click **Add**. The window should now appear as in Figure 1-6. If there were more than two values you would continue in this way until the labels for all possible values had been entered. Be sure to click **Add** after the last value label has been entered.
5. Click **OK**.

To view the labels in the "Data View" spreadsheet, click **View…Value Labels** so that a check mark ✔ appears by it. To have the numerical values appear, repeat this process to remove the check mark.

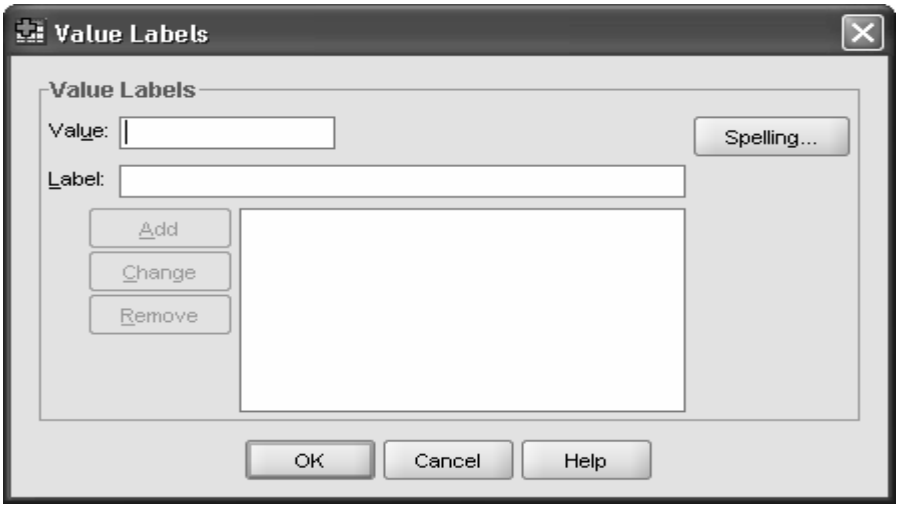


Figure 1-6

1.10 Printing in SPSS

To print graphs or tables from the SPSS output window, it is easiest to copy these into a word processor such as Word. They can then be included as part of a report. See the next section for copying SPSS output to Word.

To print out the data spreadsheet in the SPSS Data Editor, follow these steps.

1. Click **File**, then click **Print**. The "Print Untitled" window appears.
2. SPSS assumes that you want to print the entire data set. If you wish to print only a subset of the data (e.g., only one variable), you must first select the subset in the Data Editor (click on a column heading to highlight one variable, click and drag to highlight several variables, click and drag to highlight a subset of cells).
3. Click **OK**.

1.11 Copying from SPSS into Microsoft Word

There are a variety of ways to copy SPSS output into Word depending on the version of Word and the version of SPSS you have. The following methods for copying charts and tables in SPSS into Word work for all versions of Word (95, 97, 2000, and 2002) and SPSS. A chart is any picture in SPSS, such as a histogram or scatterplot, usually the result of a command in the **Graphs** menu. A table is a formatted table of values usually from commands in the **Analyze** menu.

To copy a chart or table from SPSS into Word, click on the chart in the SPSS output window to select it, then click **Edit…Copy**.

If you are using an older version of Word, you will need to use the **Copy** option for charts and the **Copy Objects** option for tables; otherwise, the figures will not copy correctly into Word.

Then go into Word, place the cursor where you want to place the chart or table and click Edit…Paste.(or click ). You can resize the chart or table in Word by clicking on it so that little "handles" appear around the edges. Click on the handle in the lower right hand corner and hold the mouse button down; a double-ended arrow should appear. While still holding the mouse

button down, you can drag the handle to make the picture bigger or smaller. If you want to move the chart or table within Word, right-click on the chart or table and select **Format Picture**. The **Format Picture** window will appear as shown in Figure 1-7. Select "Layout" from the list of options at the top of this window, select the "Behind text" wrapping style, and click OK. You can now click on the chart or table and drag it to the desired location in the Word document.

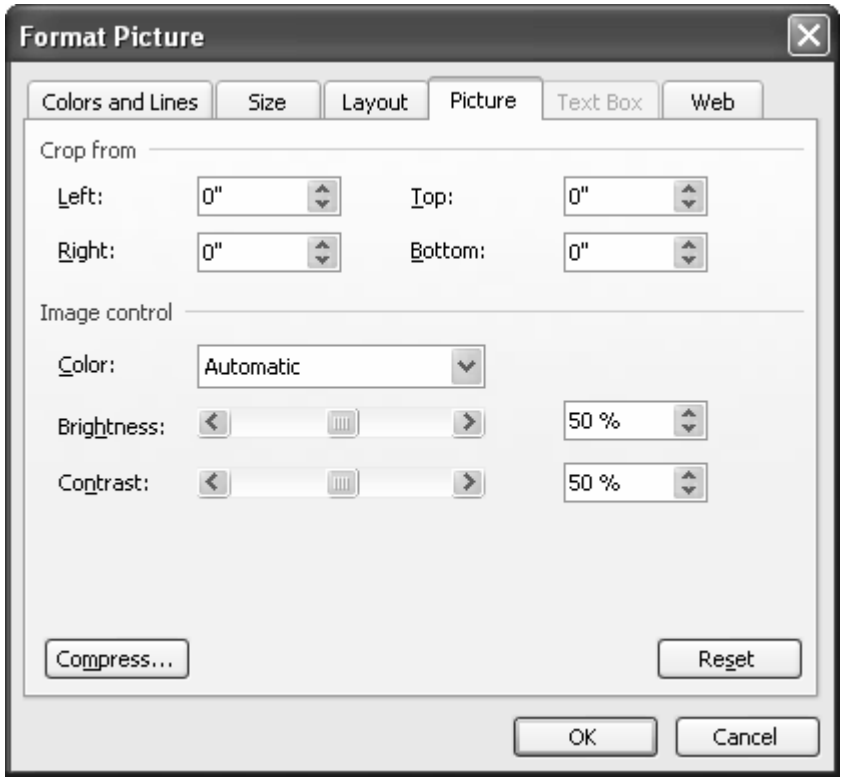


Figure 1-7

1.12 Using SPSS Help

SPSS has extensive on-line help. Click **Help** and then click **Topics**. You can select "Contents", "Index", "Search", or "Favorites" by using the index tabs at the top of this window. In "Contents", you can view a hierarchical list of topics on which you can get help. Double-click on headings to see subtopics under each topic. In "Index" you can look up specific words. In "Search" you can do a more complete search for specific words or phrases. In "Favorites" you can select previous favorite help requests. To exit "Help", click **Cancel**.

1.13 Note on E notation for numbers

Sometimes, a number like 8.750E-02 will appear in the output. This is scientific notation; it means 8.750×10^{-2} or .0875. This is a useful way of representing very small or large numbers where there wouldn't be room in the output for all the digits. The number after the E represents the power of 10 the number should be multiplied by. Equivalently, just think of this number as how many places to the left or right the decimal point should be moved; negative means to the left and positive to the right. For example, 1.343E-05 is .00001343 and 4.602E08 is 460,200,000.

1.14 Creating new variables by data transformations

Often, you will want to modify existing data to some other form and create a new variable with these ``transformed data." For example, in **Example 1-1: Student Measurements** (page 5), suppose that we would like to create a new variable which is height in inches instead of centimeters. To convert centimeters to inches, we divide by 2.54. To create this new variable in SPSS, make sure you are in the Data Editor and follow these steps.

1. Click **Transform** and click **Compute Variable**. This will open the "Compute Variable" window shown in Figure 1-8, except that the "Target Variable" and "Numeric Expression" boxes will be blank.
2. In the "Target Variable" box, type the name of the new variable. In Figure 1-8, the new variable has been named ***heighti*** (to indicate height in inches).
3. In the "Numeric Expression" box, type in the definition of the new variable. In this case, ***heighti*** is defined to be ***height/2.54***. The latter expression can either be typed in or you can use the list of variables on the lower left and the button to automatically enter the variable name ***height*** in the "Numeric Expression" box. To do so, click on the ***height*** variable name to highlight it and then click . You can then finish typing the expression.
4. Click **OK**.

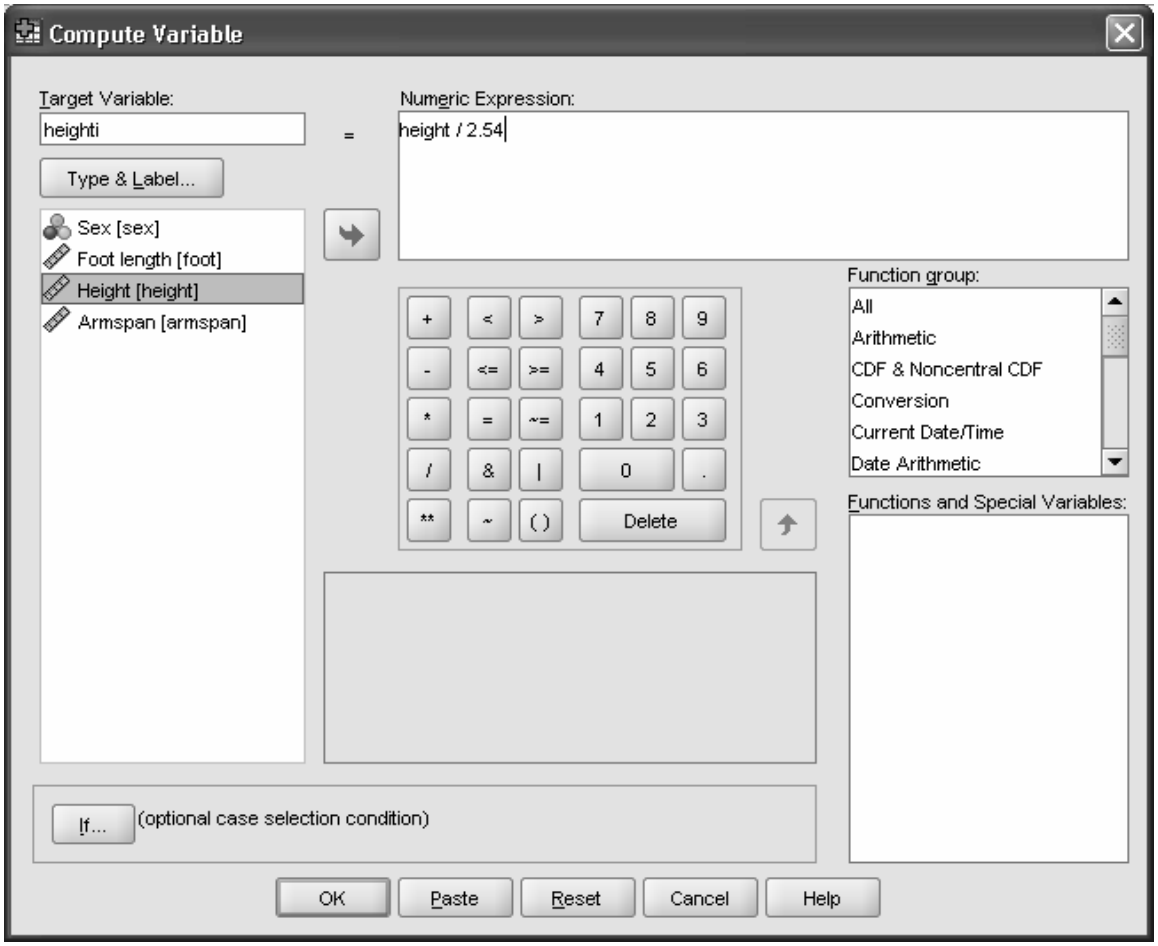


Figure 1-8

You can enter quite complex expressions in the "Numeric Expression" box involving more than one variable (for example, the ratio of two existing variables, or the natural logarithm of a variable) and using the function group and corresponding list of functions in the lower window on the right-hand side of the "Compute Variable" window. You can also use the calculator keypad to enter symbols and numbers, if desired.

As an example, computing the natural logarithm of the height variable is carried out by the following steps:

1. Select "Arithmetic" from the list of function groups. This will result in a list of functions appearing in the "Functions and Special Variables" window beneath the function groups, as shown in Figure 1-9
2. Select "Ln" from the list of functions, and click the button to move this function into the "Numeric Expression" window. The function will appear as: "LN(?)" with the question mark highlighted in blue.
3. Select ***height*** from the list of variables on the left of the screen, and click the button to replace the question mark with the ***height*** variable.
4. In the "Target Variable" box, type the name of the new variable. The new variable has been named ***logheight***. The final window appears in Figure 1-9.
5. Click **OK**.

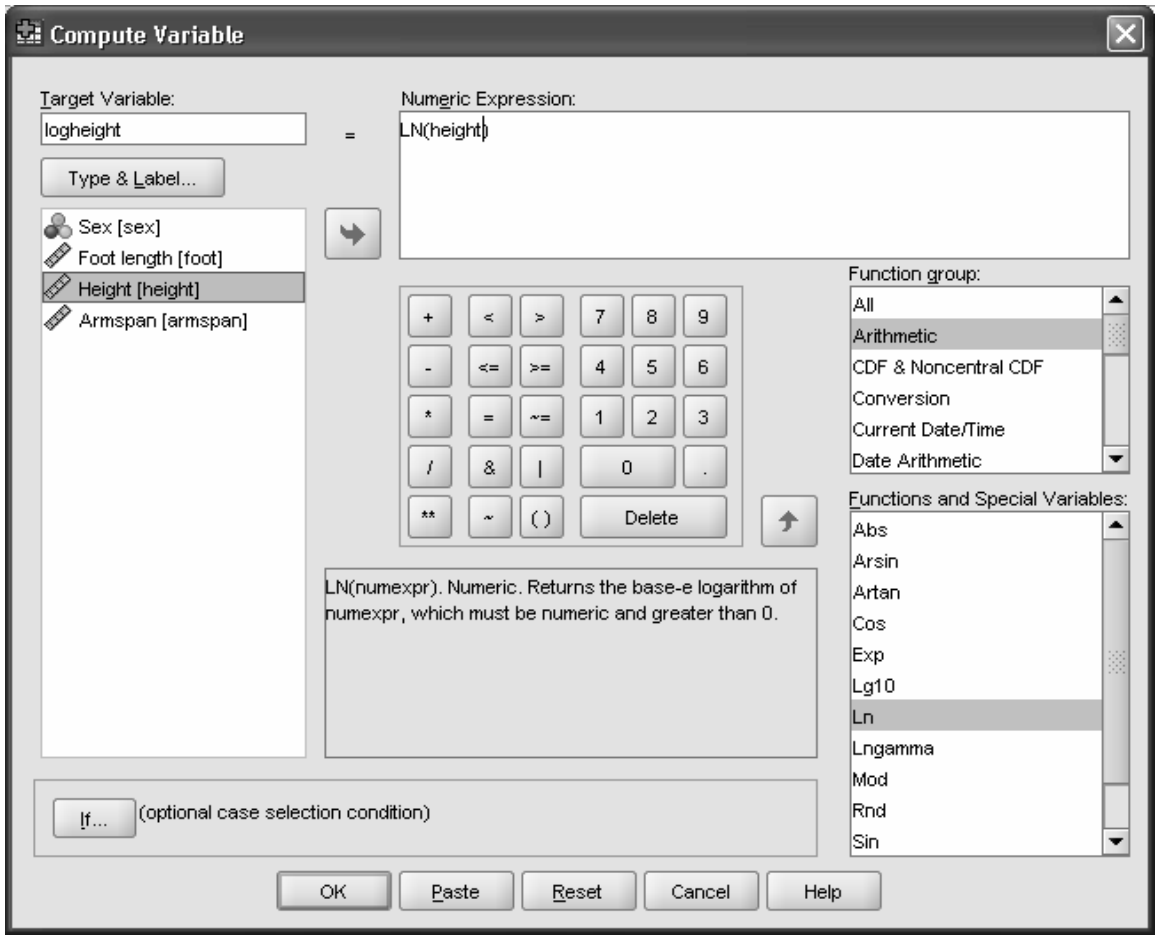


Figure 1-9

1.15 Using subsets of the data

Occasionally, you may wish to analyze only a subset of the cases in a data set. In **Example 1-1: Student Measurements.,** page 5, you may wish to examine only the females, or only people under 170 cm. in height. Temporary subsetting is accomplished by the following steps.

1. Click **Data** and click **Select Cases**. The window in Figure 1-10 appears.

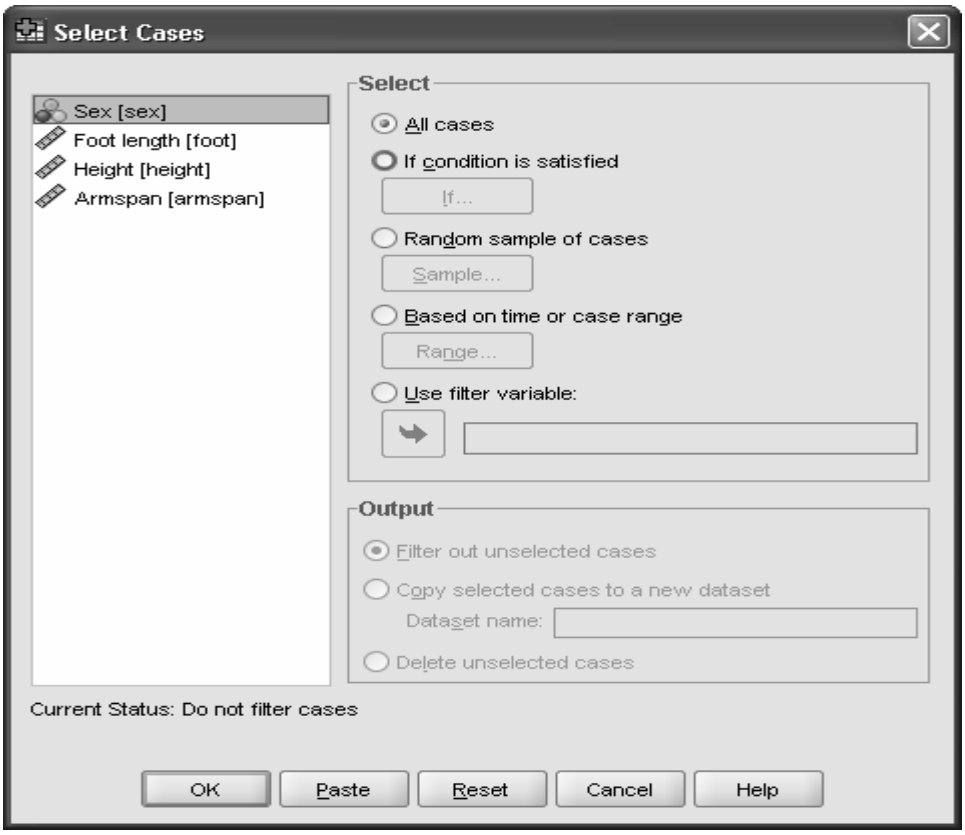


Figure 1-10

2. Click "If condition is satisfied" under "Select" so that a bullet appears by this option.
3. Click the "If…"button under the "If condition is satisfied" button. The window in Figure 1-11 appears.
4. At this point you can create any condition for selecting cases that you desire in the blank window on the right. For example, to select only females, type **sex = 0** in that box (you can also use the  button to move a variable name into the blank box instead of typing the name). To select only people with height less than 170 cm., type **height < 170** in the box. Conditions can also be combined using **&** for "and" and **|** for "or". For example, to select only females who are less than 170 cm. in height, type **sex = 0 & height < 170** in the box.
5. Once you have typed the condition for selecting cases, click **Continue**. The window in Figure 1-10 appears again. Be sure that the "Filtered" option (the default) is selected under the "Unselected Cases Are" heading near the bottom of the window. This means that the cases not satisfying the condition are not deleted from the data set and can be restored later. If the "Deleted" option is selected, cases not satisfying the condition are deleted from the data set.
6. Click **OK**.

The data set now appears as in Figure 1-12 (the condition used was **sex=0** in this example). There are slashes (/) through the case numbers of cases not selected. In addition, a new variable, ***filter_$*** is created which is 1 for selected cases and 0 for non-selected cases. At this point, you can carry out any SPSS procedures (e.g., histograms, t-tests, etc.) as you normally would, but only the selected cases will be used in the analyses.

To return to using the full data set, click **Data**, click **Select Cases**, and click on the "All cases" option under "Select" so that a bullet appears by this option. Then click **OK**.

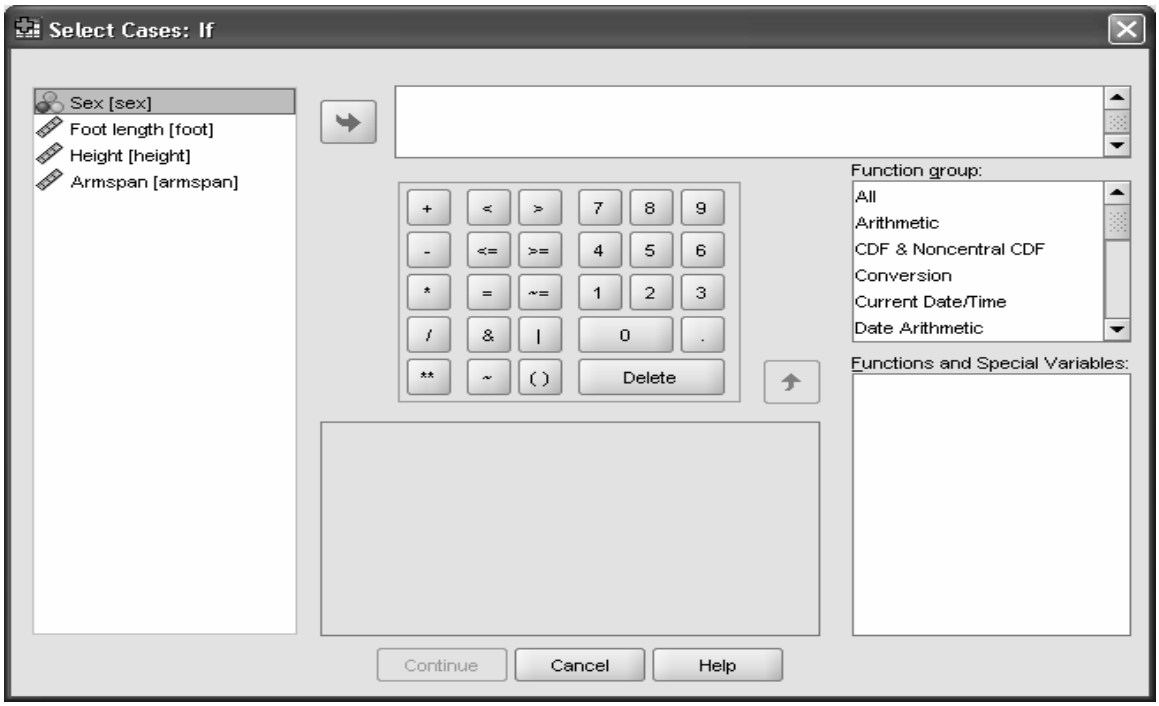
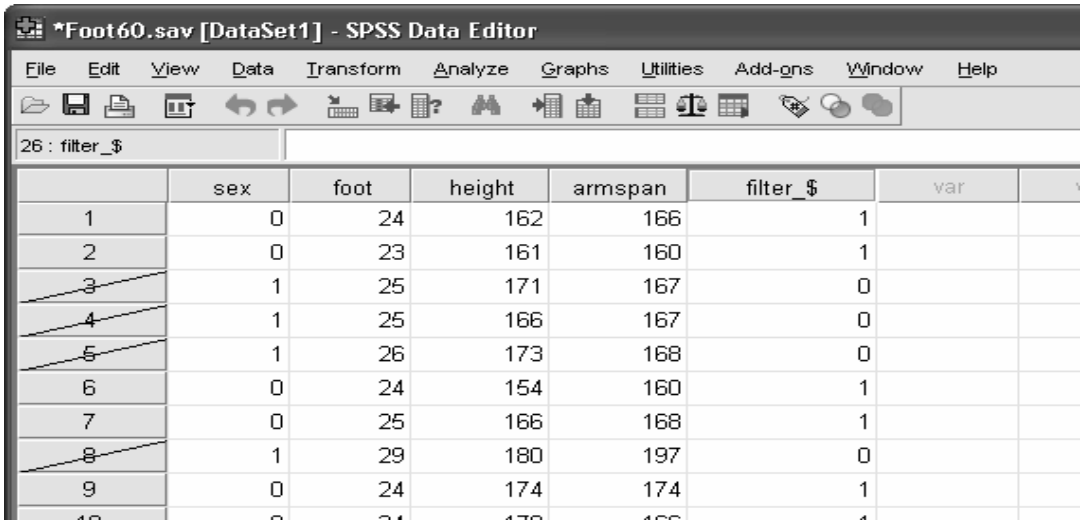


Figure 1-11



	sex	foot	height	armspan	filter_$	var
1	0	24	162	166	1	
2	0	23	161	160	1	
~~3~~	1	25	171	167	0	
~~4~~	1	25	166	167	0	
~~5~~	1	26	173	168	0	
6	0	24	154	160	1	
7	0	25	166	168	1	
~~8~~	1	29	180	197	0	
9	0	24	174	174	1	

Figure 1-12