

SPSS: Getting Started

For Windows



Updated: August 2012

Table of Contents

Section 1: Overview.....	3
1.1 Introduction to SPSS Tutorials	3
1.2 Introduction to SPSS.....	3
1.3 Overview of SPSS for Windows.....	3
Section 2: Entering Data in SPSS	4
2.1 Starting SPSS	4
2.2 The Data Editor.....	4
2.3 The Syntax Editor	5
2.4 The Output Viewer	6
2.5 Importing Data from Excel Files	8
2.6 Importing data from ASCII files.....	10
Section 3: Creating and Modifying Data in SPSS	11
3.1 Creating and Defining Variables	11
3.2 Inserting and Deleting Cases and Variables	13
3.3 Computing New Variables.....	13
3.4 Recoding Variables.....	16
3.5 Sorting Cases	18
3.6 Selecting Cases	18
3.7 Listing Cases.....	20
Conclusion	22

Section 1: Overview

1.1 Introduction to SPSS Tutorials

This document is the first of a series of four modules intended for beginning SPSS users, providing an overview of SPSS for Windows. This first module introduces readers to the SPSS for Windows environment, and discusses how to create or import a dataset, transform variables, manipulate data, and perform descriptive statistics. The second module describes some commonly used inferential statistics, the third module discusses graphical display of output, and the fourth module covers other advanced topics. All modules can be found at <https://stat.utexas.edu/training/software-tutorials>. Throughout these modules, a single dataset, Employee data.sav, is used for all examples. This example dataset is provided with recent versions of SPSS. Thus, you will have access to the dataset and will be able to use SPSS to test your knowledge by replicating the examples contained in this document. Although the present documentation assumes SPSS Version 14, it will still be useful to users of SPSS on the Macintosh platform as well as many earlier, similar versions of SPSS. If you are a University of Texas affiliate and do not have access to SPSS or would like the software for your personal computer, visit the Software Distribution Services Web page at <http://www.utexas.edu/its/sds/> to get more information about obtaining the latest version of SPSS.

1.2 Introduction to SPSS

SPSS is a software package used for conducting statistical analyses, manipulating data, and generating tables and graphs that summarize data. Statistical analyses range from basic descriptive statistics, such as averages and frequencies, to advanced inferential statistics, such as regression models, analysis of variance, and factor analysis. SPSS also contains several tools for manipulating data, including functions for recoding data and computing new variables, as well as for merging and aggregating datasets. SPSS also has a number of ways to summarize and display data in the form of tables and graphs.

1.3 Overview of SPSS for Windows

SPSS for Windows consists of five different windows, each of which is associated with a particular SPSS file type. This document discusses the two windows most frequently used in analyzing data in SPSS, the Data Editor and the Output Viewer windows. In addition, the Syntax Editor and the use of SPSS command syntax is discussed briefly. The Data Editor is the window that is open at start-up and is used to enter and store data in a spreadsheet format. The Output Viewer opens automatically when you execute an analysis or create a graph using a dialog box or command syntax to execute a procedure. The Output Viewer contains the results of all statistical analyses and graphical displays of data. The Syntax Editor is a text editor where you compose SPSS commands and submit them to the SPSS processor. All output from these commands will appear in the Output Viewer. This document focuses on the methods necessary for inputting, defining, and organizing data in SPSS.

Section 2: Entering Data in SPSS

2.1 Starting SPSS

To start SPSS, go to the Start icon on your Windows computer. You should find an SPSS icon under the Programs menu item. You can also start SPSS by double-clicking on an SPSS file. When the program opens, it will present you with a “What would you like to do?” dialog box. For now, hit “Cancel” to dismiss the box.

2.2 The Data Editor

The Data Editor window displays the contents of the working dataset. It is arranged in a spreadsheet format that contains variables in columns and cases in rows. There are two sheets in the window. The *Data View* is the sheet that is visible when you first open the Data Editor; this sheet contains the data. You can access the second sheet by clicking on the tab labeled *Variable View*. While the second sheet is similar in appearance to the first, it does not actually contain data. Instead, this second sheet contains information about the variables in the dataset. Beginning with version 14, you can have multiple datasets open at one time in the Data Editor (however, this can be confusing, and we recommend keeping only one dataset open at a time while you are first getting familiar with the program.) Datasets that are currently open are called *working datasets*; all data manipulations, statistical functions, and other SPSS procedures operate on these datasets. Data can be directly entered in SPSS, or a file containing data can be opened in the Data Editor. From the menu in the Data Editor window, choose the following menu options:

File
Open
Data...

The *Open File* dialog box should automatically open to the SPSS directory of example files. Choose *Employee data.sav* from the list and click **Open**. Your Data Editor should now look like this:

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority
1	1	m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0
2	2	m	05/23/1958	16	1	\$40,200	\$18,750	98	36	0
3	3	f	07/26/1929	12	1	\$21,450	\$12,000	98	381	0
4	4	f	04/15/1947	8	1	\$21,900	\$13,200	98	190	0
5	5	m	02/09/1955	15	1	\$45,000	\$21,000	98	138	0
6	6	m	08/22/1958	15	1	\$32,100	\$13,500	98	67	0
7	7	m	04/26/1956	15	1	\$36,000	\$18,750	98	114	0

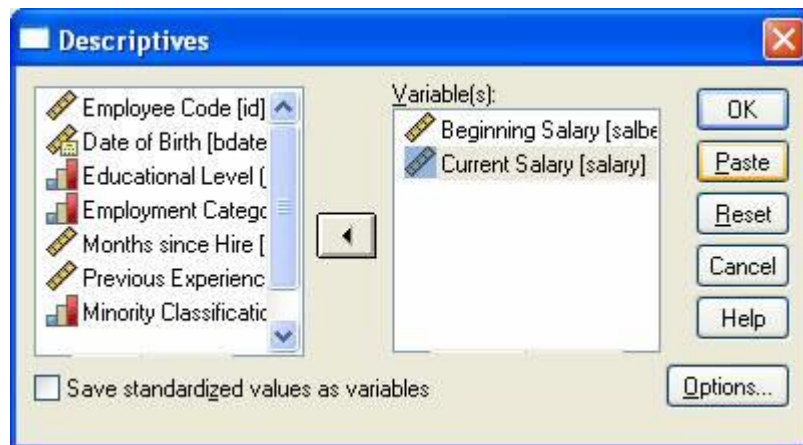
If the file you want to open is not an SPSS data file, you can often use the *Open* menu item to import that file directly into the Data Editor. If a data file is not in a format that SPSS recognizes,

then try using the software package in which the file was originally created to translate it into a format that can be imported into SPSS (e.g., Excel).

2.3 The Syntax Editor

Another important window in the SPSS environment is the Syntax Editor. In earlier versions of SPSS, all of the procedures performed by SPSS were submitted through the use of syntax, which instructed SPSS on how to process your data. More recent versions contain pull-down menus with dialog boxes that allow you to submit commands to SPSS without writing syntax. This SPSS for Windows tutorial focuses on the use of dialog boxes to execute procedures; however, there are at least two reasons why you should be aware of SPSS syntax, even if you plan to primarily use the dialog boxes. First, not all procedures are available through the dialog boxes. Therefore, you may occasionally have to submit commands from the Syntax Editor. Second, the Syntax Editor is a useful way to save a log of what you have done, and to re-run what you have done at a later date. The dialog boxes available through the pull-down menus have a button labeled **Paste**, which will print the syntax for the procedure you are running in the dialog box environment to the Syntax Editor. Thus, you can easily generate SPSS syntax without typing in the Syntax Editor. This process is illustrated below.

The following dialog box is used to generate descriptive statistics. (You can get this dialog box by choosing **Analyze**, then **Descriptive Statistics**, then **Descriptives**, then clicking over the two variables using the arrow button.)



By clicking on the **Paste** button, the procedure that the above dialog box is prepared to run will be written in the form of SPSS syntax to the Syntax Editor. Thus, clicking the **Paste** button in the above example would produce the following syntax:

DESCRIPTIVES

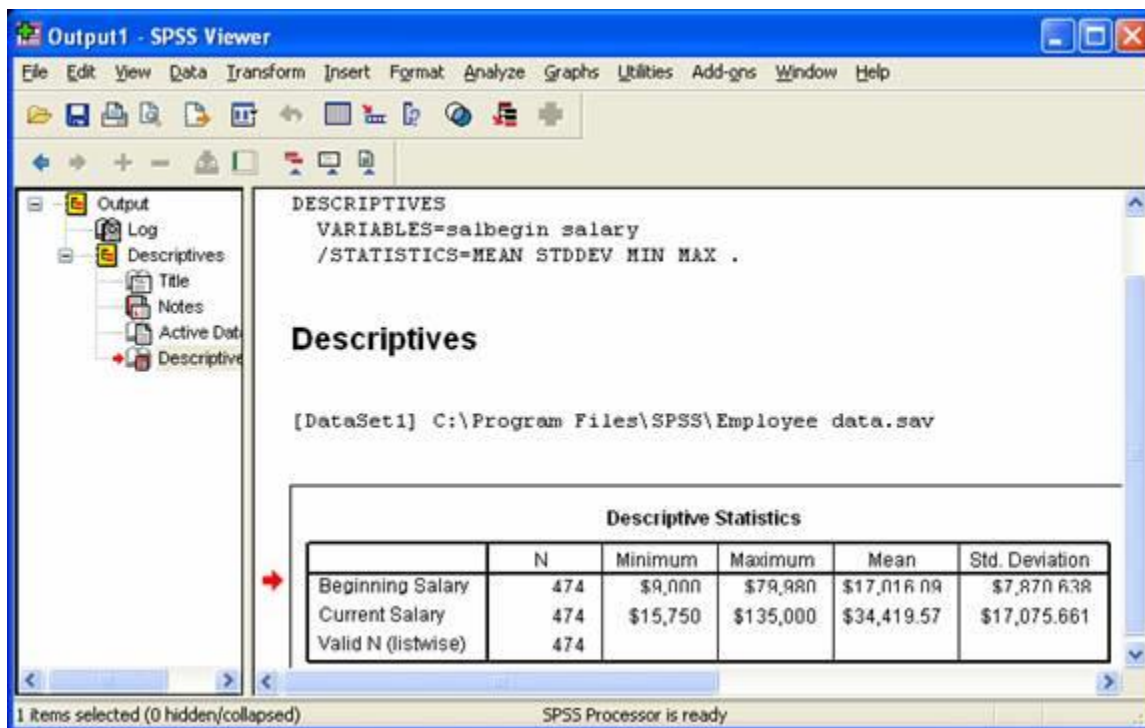
VARIABLES=salbegin salary

/STATISTICS=MEAN STDDEV MIN MAX .

This syntax will produce exactly the same output as would be generated by clicking the **OK** button in the above dialog box. The syntax that is printed to the Syntax Editor can then be saved and run at a later time, as long as the same dataset (or at least a dataset containing the variables with the same names) is active in the Data Editor window. Saving syntax is useful if you think you may want to rerun your analysis after you add more data, or if you want to run the same analysis on another dataset that contains the same variables.

2.4 The Output Viewer

When you execute a command for a statistical analysis, regardless of whether you used syntax or dialog boxes, the output will be printed in the Output Viewer. An example of the output viewer is shown below:



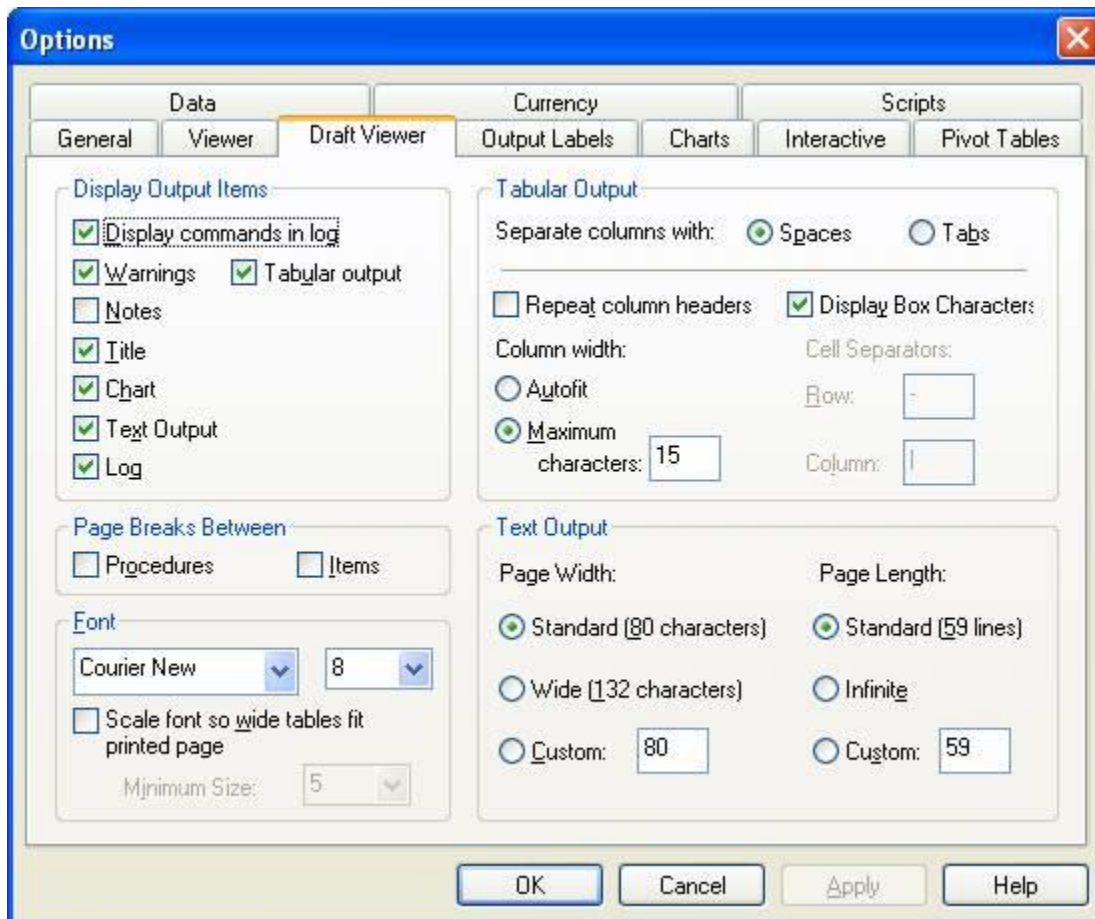
The left frame of the Output Viewer contains an outline of the objects contained in the window. For example, the icon labeled *Log* represents the command syntax shown at the top of the figure. Everything under *Descriptives* in the outline refers to objects associated with the descriptive statistics. The *Title* object refers to the bold title *Descriptives* in the output. The *Active Dataset* object refers to the line in the output that designates which dataset was used to run the analysis. The highlighted icon labeled *Descriptive Statistics* refers to the table containing descriptive statistics. The *Notes* icon has no referent in the above example, but it would refer to any notes that appeared between the title and the table. This outline can be useful for navigating in your Output Viewer when you have large amounts of output. By clicking on an icon, you can move to the location of the output represented by that icon in the Output Viewer. You can also copy,

paste, or delete objects by first highlighting them in the outline and then performing the operation you want.

You can control what is displayed in your output by using the *Options* menu item on the *Edit* menu:

Edit Options...

Selecting this option will produce the following dialog box:



This figure shows the *Options* dialog box with the *Draft Viewer* tab selected, to choose which options you want to appear in the Output Viewer. Most commands are selected by default. Here, the *Display commands in log* option, normally unselected, was selected so that the command syntax will be written to the log in the Output Viewer. This can be useful for keeping track of which procedures you have executed.

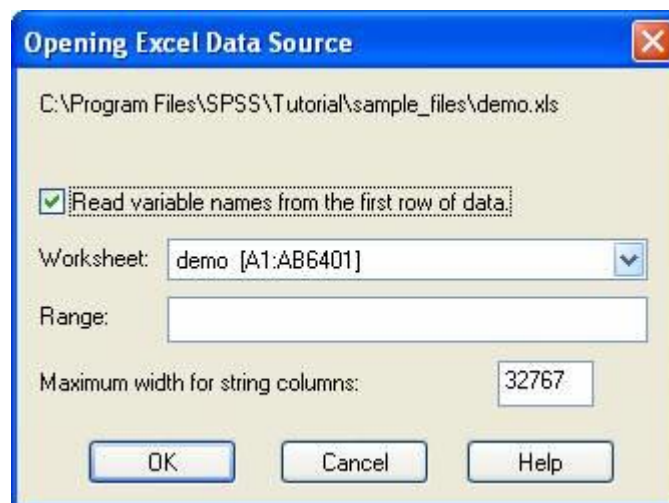
2.5 Importing Data from Excel Files

Data can be imported into SPSS from Microsoft Excel and several other applications with relative ease. This document describes a method for importing an Excel spreadsheet into SPSS. If you are working with a spreadsheet in another software package, you may want to save your data as an Excel file, then import it into SPSS. If you have a spreadsheet that is arranged in a database format (e.g., you have several tables in your Workbook that are related through identification fields), there is another method for importing Excel file that you might consider that will merge tables within your database as part of the import procedure. It is described in our online tutorial of [SPSS Data Manipulation and Advanced Topics](#).

In order to easily import Excel data into SPSS, make sure your Excel spreadsheet is formatted as follows: (1) the spreadsheet should have a single row of variable names across the top of the file, and each variable name should begin with ordinary letters, rather than with any special characters, (2) the data should begin in the first column and second row of the Excel file, and (3) any graphs, labels, or extra text that is not part of the dataset should be deleted. To open an Excel file, select the following options from the menu in the Data Editor window in SPSS:

File
Open
Data...

First, select the desired location on disk using the *Look in* option. Next, select Excel from the *Files of type* drop-down menu. The file you want should now appear in the main box in the *Open File* dialog box. You can open it by double-clicking on it. You will be presented with one more dialog box:



Your Excel spreadsheet should have all variable names on the top row, so leave the *Read variable names* option checked. Since Excel files can consist of multiple worksheets, the *Worksheet* drop-down menu allows you to choose which worksheet you wish to open. You may ignore the remaining options and choose **OK**. You should now see data in the Data Editor

window. Check to make sure that all variables and cases were read correctly. Next, save your dataset in SPSS format by choosing the **Save** option in the **File** menu.

2.6 Importing data from ASCII files

Data are often stored in an ASCII file format, alternatively known as a text or flat file format. Typically, columns of data in an ASCII file are separated by a space, tab, comma, or some other character. To open these files, choose:

File
Read Text Data

The Text Import Wizard will first prompt you to select a file to import. After you have selected a file, you will go through a series of dialog boxes that will provide you with several options for importing data. Once you have imported your data and checked it for accuracy, be sure to save a copy of the dataset in SPSS format by selecting the *Save* or *Save As* options from the *File* menu:

File
Save
Save As...

Section 3: Creating and Modifying Data in SPSS

3.1 Creating and Defining Variables

After data are in the Data Editor window, there are several things that you may want to do to describe your data. Before describing the process for defining variables, an important distinction should be made between two terms that are often confused: *variable* and *value*. A variable is a measure or classification scheme that can have several values. Values are the numbers or categorical classification representing individual instances of the variable being measured. For example, a variable could be created for employment classification status. Each individual in the dataset would be assigned a value representing their job classification. For instance, we could assign clerks the value 1, custodial workers the value 2, and managers the value 3.

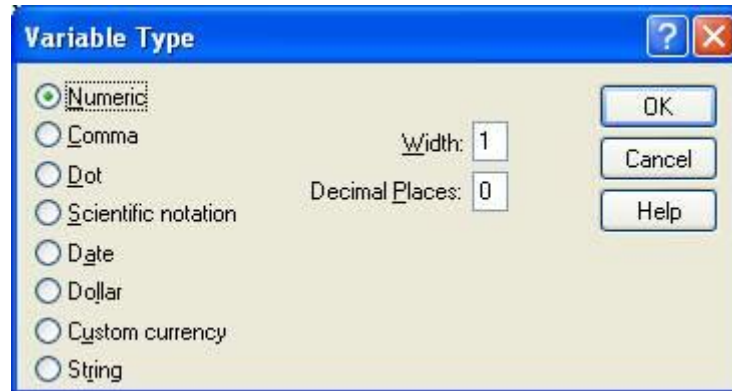
One reason to define information about your variables is to help you interpret the output. For example, if your employment categories are coded as either 1, 2, or 3, it may be unwieldy to read the output if you are constantly trying to remember which number represents which categories. One advantage of defining variables is that these values can be assigned labels that will appear in your output, thus making it much easier to interpret. Another aspect of defining variable information is to provide SPSS with information about the type of data in your dataset, which is often critical for SPSS to correctly process analyses.

You can define information about your variables by clicking the *Variable View* tab. Doing so will bring the *Variable Information* sheet to the foreground. You can also access this sheet by double-clicking one of the variable names at the top of the columns in the Data Editor. The advantage of the second method is that it takes you to the row for the variable whose column head you clicked. Using the *Employee data.sav* dataset, you will see a spreadsheet organized as the one below:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	4	0	Employee Cod	None	None	8	Right	Scale
2	gender	String	1	0	Gender	{f, Female}...	None	1	Left	Nominal
3	bdate	Date	10	0	Date of Birth	None	None	13	Right	Scale
4	educ	Numeric	2	0	Educational Le	{0, 0 (Missing)}	0	8	Right	Ordinal
5	jobcat	Numeric	1	0	Employment C	{0, 0 (Missing)}	0	8	Right	Ordinal
6	salary	Dollar	8	0	Current Salary	{\$0, missing}...	\$0	8	Right	Scale
7	salbegin	Dollar	8	0	Beginning Sala	{\$0, missing}...	\$0	8	Right	Scale
8	jobtime	Numeric	2	0	Months since	{0, missing}...	0	8	Right	Scale
9	prevexp	Numeric	6	0	Previous Exper	{0, missing}...	None	8	Right	Scale
10	minority	Numeric	1	0	Minority Classi	{0, No}...	9	8	Right	Ordinal

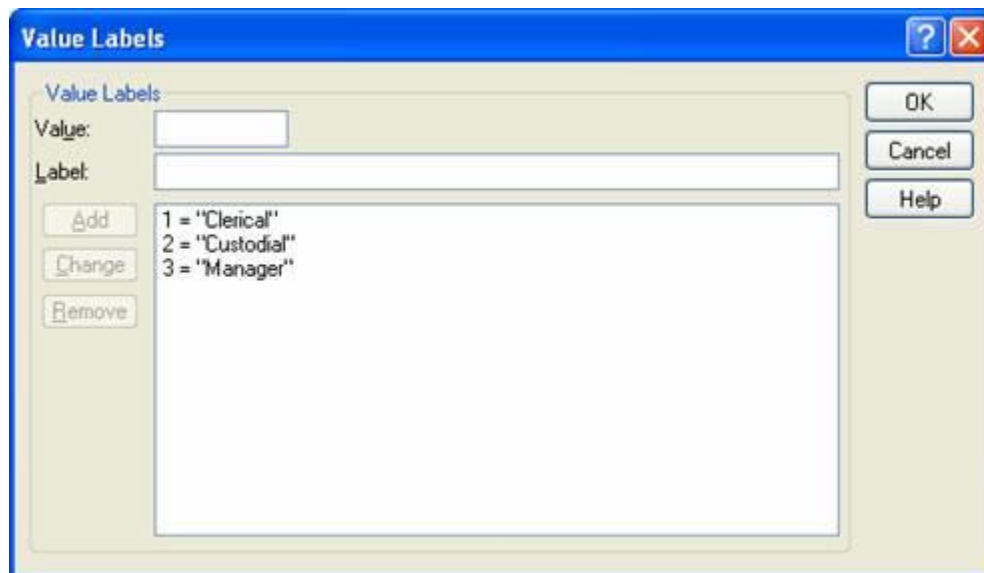
Many of the cells in the spreadsheet contain hidden dialog boxes that can be activated by clicking on a cell. If you see a gray box appear on the right side of the cell when you first click on the cell, this indicates that there is a hidden dialog box which can be accessed by clicking on

that box. For example, clicking on the box in the cell for the *Type* column for the variable *jobcat* produces the following dialog box:



This box allows you to define the type of data for variables. For example, you will be presented with *Numeric*, *String*, and *Date* options. Thus, if you wanted to define *jobcat* as a string variable rather than the default numeric variable type, you would choose the *String* option.

Looking back at the Variable View, the *Missing Values* column allows you to define which values of a variable should be treated as missing data. The *Label* column is used to define labels for variables. The *Values* column is used to assign labels to the particular values of a variable. For example, the following dialog box shows that the *jobcat* variable that has been assigned the values 1, 2, and 3 for the labels *Clerical*, *Custodial*, and *Manager*.



To define variables as shown above, you should first enter the value (e.g., 1) in the box labeled *Value*, then enter the label associated with that value (e.g., *Clerical*), and click on the **Add** button. Repeat this process for each value you want to label.

3.2 Inserting and Deleting Cases and Variables

You may want to add new variables or cases to an existing dataset. For example, you may want to add data about participants' ages to an existing dataset. To insert a new variable, go to the Data View and right-click on the variable name next to the place that you would like to insert a new variable, then choose *Insert Variables*. To insert a case, right-click the row number below the place that you would like to insert the new row, and choose *Insert Case*.

You may also want to delete cases or variables from a dataset. To do that, select a row or column, and use the **Delete** key on your keyboard to delete the highlighted area. Alternatively, you can use the *Clear* option in the *Edit* menu.

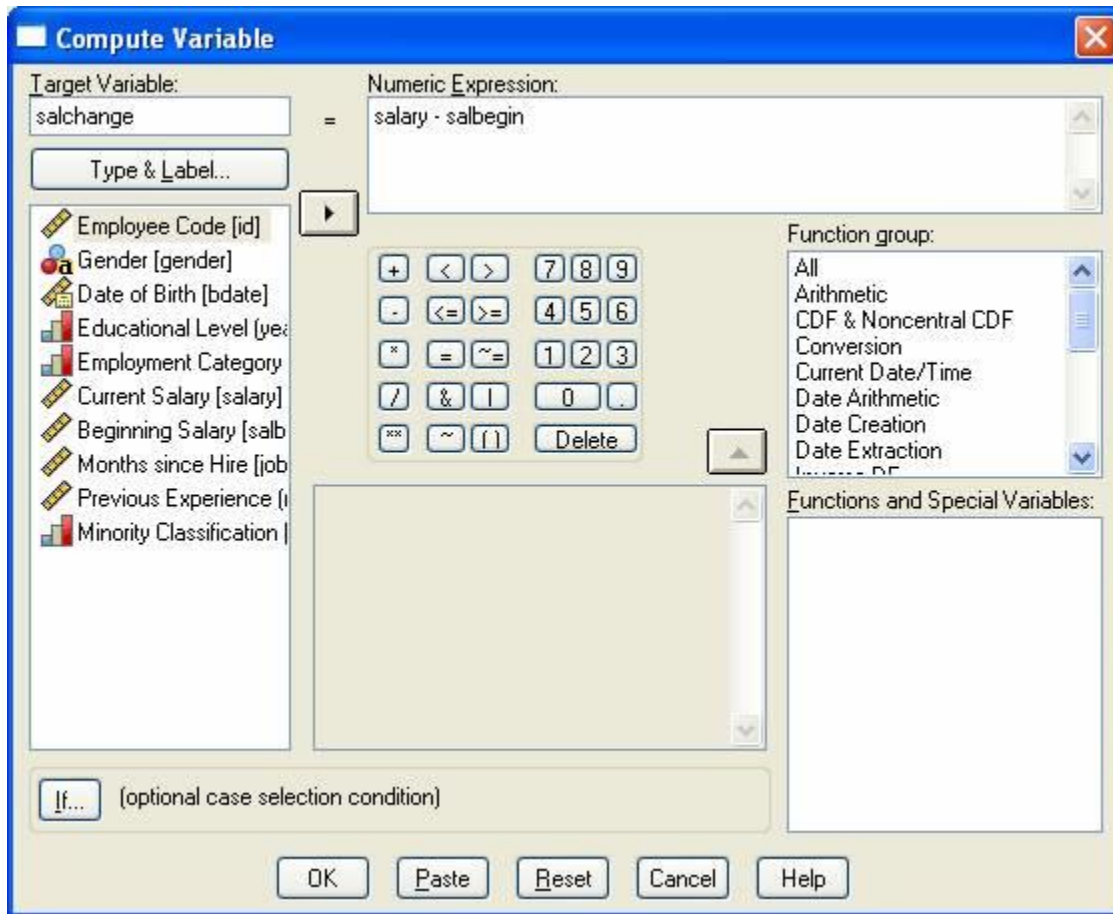
3.3 Computing New Variables

You may want to create new variables in your datasets. For example, *Employee data.sav* contains employees' salaries in terms of their beginning and current salaries, but you may also want the difference between starting salary and present salary; this new variable could be computed by subtracting the starting salary from the present salary. Alternatively, you may want to transform an existing variable. For example, the variable *jobtime* represents months of experience on the job, but you may wish to analyze data in terms of years on the job; in this case, you could compute the new variable by dividing *jobtime* by 12.

In both situations, the new variable can be created using the *Compute* option available from the menu in the Data Editor:

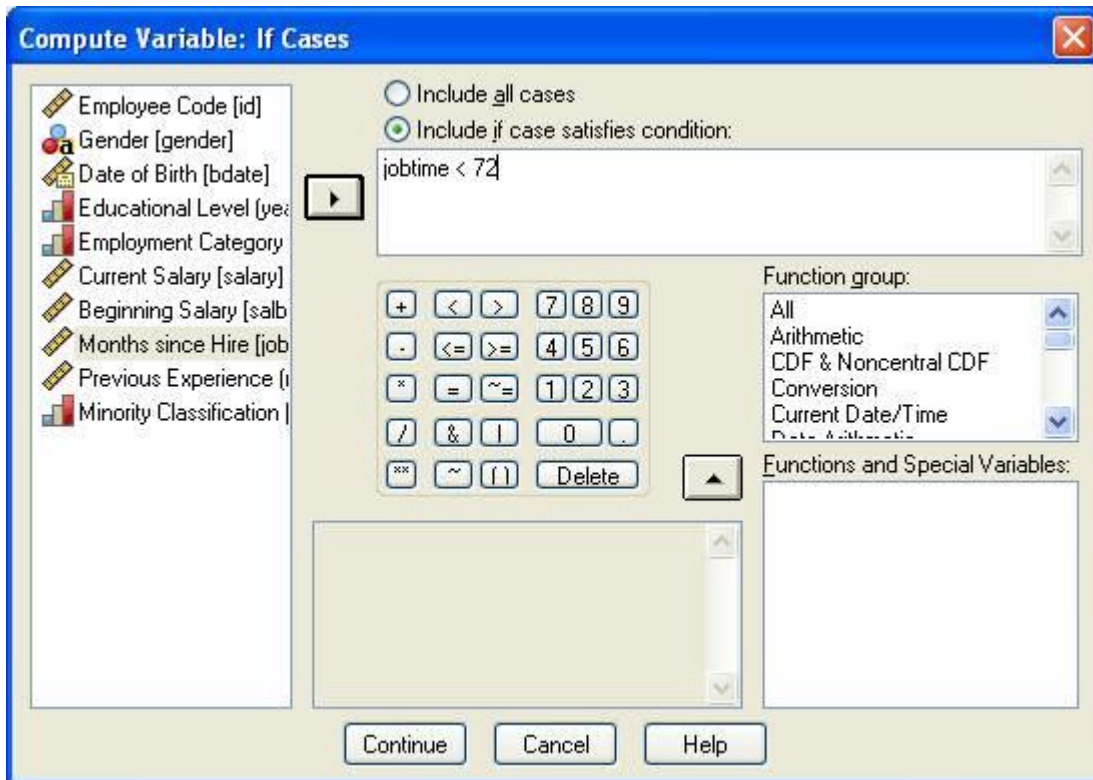
Transform
Compute...

To create a new variable, type its name in the box labeled *Target Variable*. The expression defining the variable being computed will appear in the box labeled *Numeric Expression*. This expression can either be typed into the box directly, or you can use the buttons located below the *Numeric Expression* box to input values or operators.



The example shown above demonstrates the computation of a new variable. This new variable, *salchange*, will be the difference between an employee's current salary and the employee's beginning salary. The new variable will appear in the rightmost column of the working dataset.

Variables can also be computed conditionally. For instance, if in the above example, you were only interested in the change in salaries for people who began working for the company within the last six years, you could create a condition that would compute a new variable only if an employee had begun employment within the last 72 months. To do this, first click on the button labeled **If**, which will produce the following dialog box:



First, click on the button labeled *Include if case satisfies condition* to activate the grayed-out areas of the dialog box. Then, specify the condition for computing a new variable in the box at the top right. You can either type in the condition or click on variables in the list on the left side of the dialog box and use the buttons on the bottom middle of the dialog box. Variables can be moved to the conditional box by clicking on the variable's name, then clicking the arrow button between the two boxes. Clicking on the buttons on the bottom left of the dialog box will cause the character on the button to be displayed at the location of the cursor in the input box.

The above example illustrates the definition of a condition that requires cases to have less than 72 months' experience in order to be included in the computation of the new variable. The variable *jobtime* represents the number of months since an employee has been hired. Here, only employees with *jobtime* < 72, or who have been working at the company for less than 72 months, will be included. Click the **Continue** button to return to the previous dialog box, then click **OK**. The new variable should appear in the rightmost column of your dataset. The first several rows for this variable will be blank because these people have more than 72 months of experience; scroll down about 2/3 of the way down the dataset, and you will see the values for those with less than 72 months of experience.

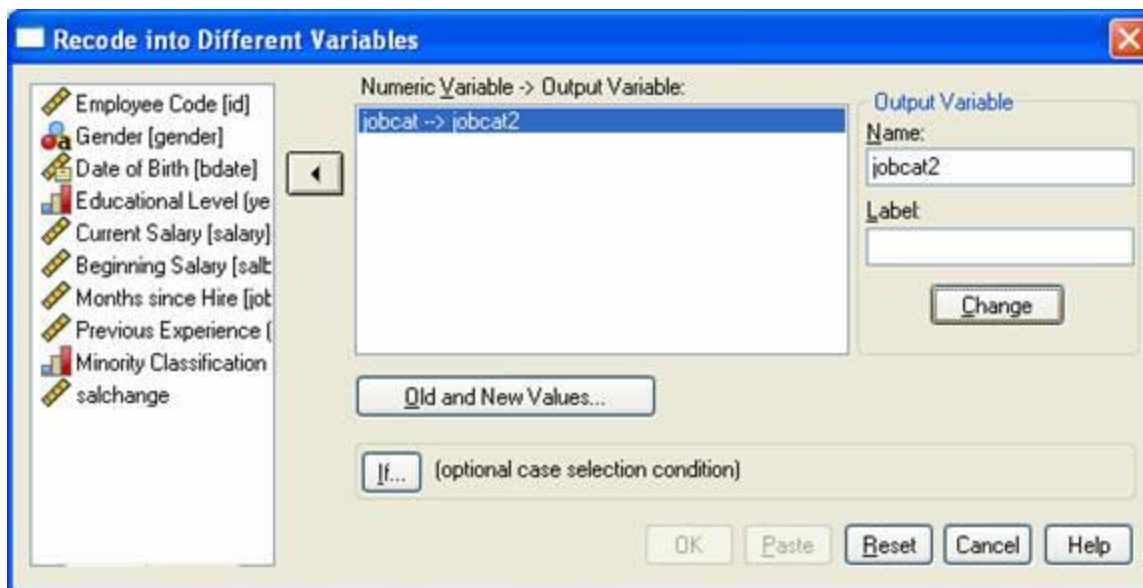
3.4 Recoding Variables

You can also modify the values of existing variables in your dataset. For example, the variable *jobcat* codes an employee's status in three categories, but for a particular analysis you may want to combine two of these classifications into a single category. To do this, use the *Recode* option from the menu in the Data Editor:

Transform Recode

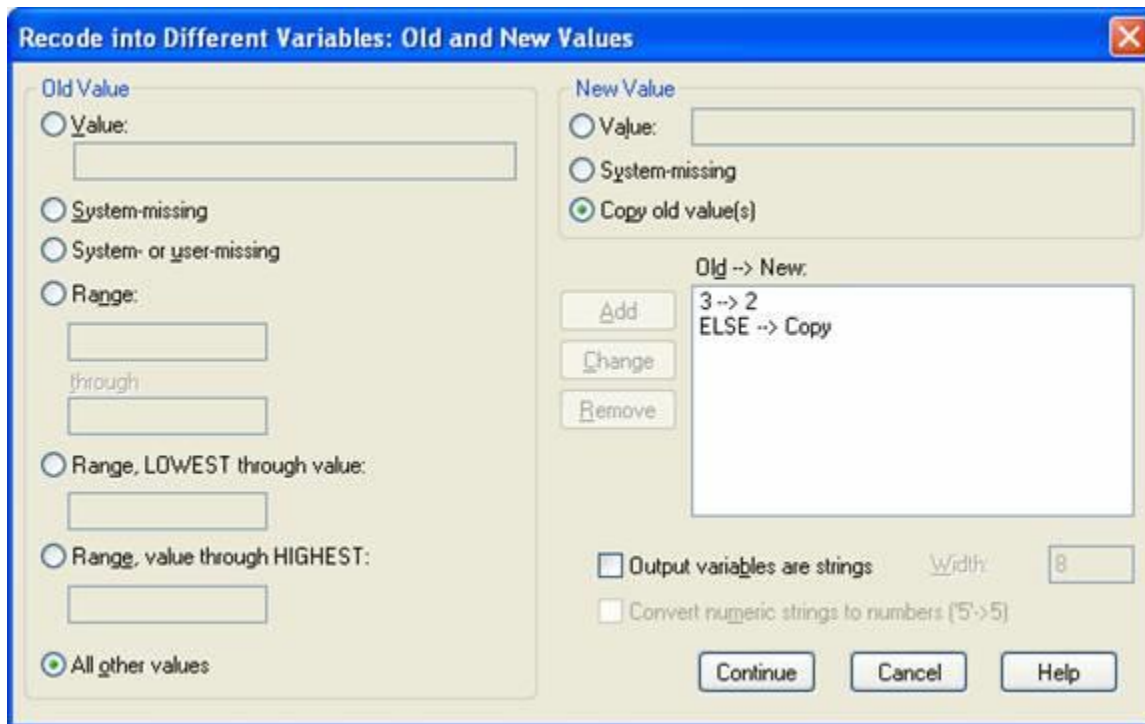
You will be offered two options for recoding variables in the *Recode* submenu. The *Into Same Variables* option changes the values of the existing variables, whereas the *Into Different Variables* option is used to create a new variable with the recoded values. Both options are essentially the same, except that recoding into a different variable requires you to supply a new variable name. We typically recommend using the *Into Different Variables* option, so that you do not overwrite your original data.

The following example illustrates the use of the *Recode Into Different Variables* option to recode *jobcat* into a new two-category variable *jobcat2*.



First, a variable from the existing dataset should be selected by clicking on that variable, then clicking the arrow button in the middle of the dialog box. This will result in the selected variable being displayed in the box labeled *Numeric Variable -> Output Variable*. Next, you must supply the name of the new variable, and optionally you can supply a label for the new variable. Then click **Change**. After a new variable name has been supplied, click on the button labeled **Old and New Values**.

The original value of the variable being recoded is entered in the box labeled *Old Value*, and the new value is entered in the box labeled *New Value*. After values are entered in these boxes, click on the button labeled **Add** to complete the recode process.



Here, the old variable *jobcat* has 3 values, and we wish to recode it into only two values. In the example dataset, *jobcat* has three values: 1, 2, and 3. If the goal were to combine cases with the values 2 and 3, this could be accomplished by recoding cases with the value 3 into 2's. Enter 3 in the box labeled *Old Value* and enter 2 in the box labeled *New Value*, then click **Add**. This can be repeated for as many of the values as necessary. In this case, we want to simply copy all the other values; under *Old Value*, click on *All other values*, and under *New Value*, click *Copy old values*, then choose **Add**. Now that all the old values have been accounted for, you may click **Continue** and then **OK**.

Values can also be recoded conditionally. The process for recoding values on the basis of a condition is essentially identical to the process for conditionally computing new variables discussed in the previous section: when you click on the **If** button in the main *Recode* dialog box, the same dialog box that was obtained from clicking **If** in the *Compute* dialog box will appear with the same options.

3.5 Sorting Cases

Sorting cases allows you to organize rows of data in ascending or descending order on the basis of one or more variable. For example, the data could be sorted by job category, so that all of the cases coded as job category 1 appear first in the dataset, followed by all of the cases that are labeled 2 and 3 respectively. The data could also be sorted by more than one variable. For example, within job category, cases could be listed in order of their salary. The *Sort Cases* option is available under the *Data* menu item in the Data Editor:

Data Sort Cases...

The dialog box that results from selecting *Sort Cases* presents only a few options:



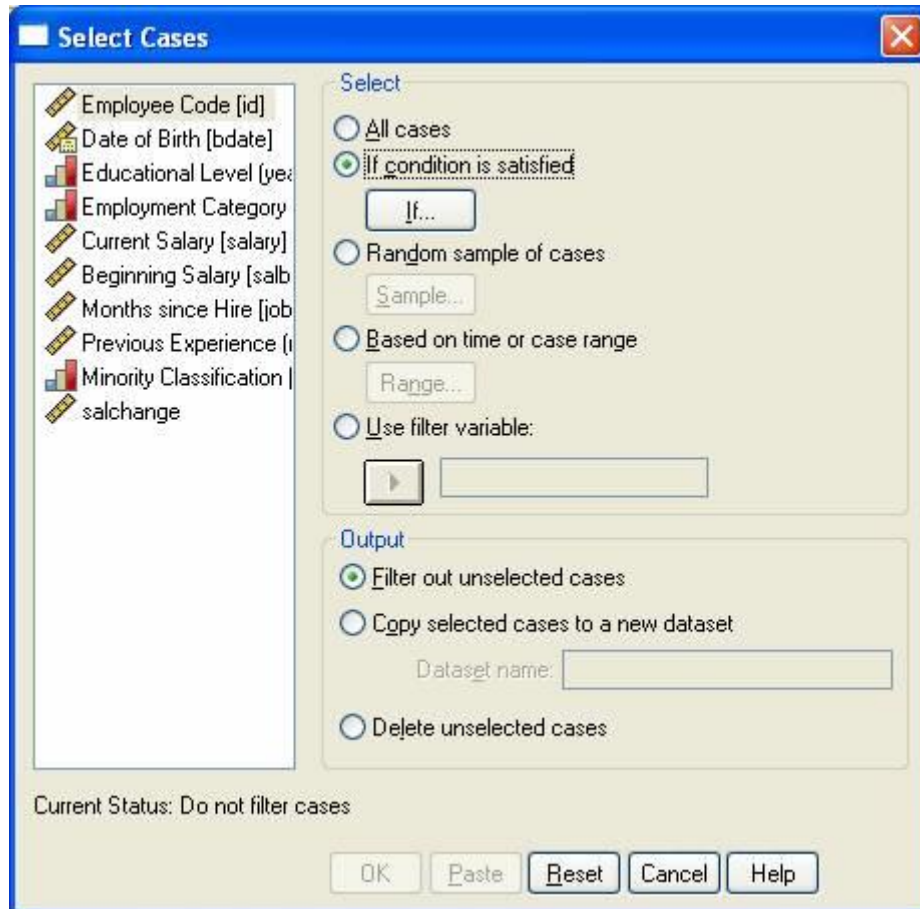
To choose whether the data are sorted in ascending or descending order, select the appropriate button. You must also specify on which variables the data are to be sorted. The hierarchy of such a sorting is determined by the order in which variables are entered in the *Sort by* box. Variables are sorted by the first variable entered, then the next variable is sorted within that first variable. Here, *jobcat* was the first variable entered, followed by *salary*; accordingly, the data would first be sorted by *jobcat*, then, within each of the job categories, data would be sorted by *salary*.

3.6 Selecting Cases

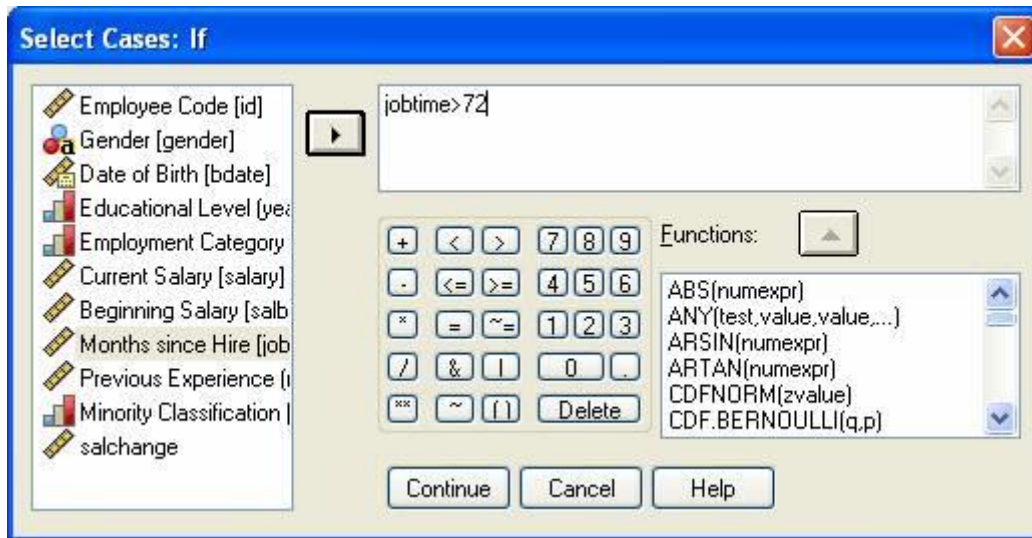
You can analyze a specific subset of your data by selecting only certain cases in which you are interested. For example, you may want to do a particular analysis on employees only if the employees have been with the company for greater than six years. This can be done by using the *Select Cases* menu option, which will either temporarily or permanently remove cases you don't want from the dataset. The *Select Cases* option is available under the *Data* menu item:

Data Select Cases...

Selecting this menu item will produce the following dialog box. This box contains a list of the variables in the active data file on the left and several options for selecting cases on the right.



Selecting one of these options will produce a second dialog box that prompts you for the particular specifications in which you are interested. For example, selecting the *If condition is satisfied* option and clicking on the **If** button (as was done in the example) results in a second dialog box, as shown below. The portion of the dialog box labeled *Output* gives you the option of temporarily or permanently removing data from the dataset. The *Filter* option will remove data from subsequent analyses until the *All Cases* option is reset, at which time all cases will again be active and used in further analyses. The *Copy* option will save the selected cases to a new dataset. The *Delete* option will remove unselected cases from the working dataset; be very careful with this option, because if the dataset is subsequently saved, these cases will be permanently deleted. Here, we have chosen to use the *Filter* option.

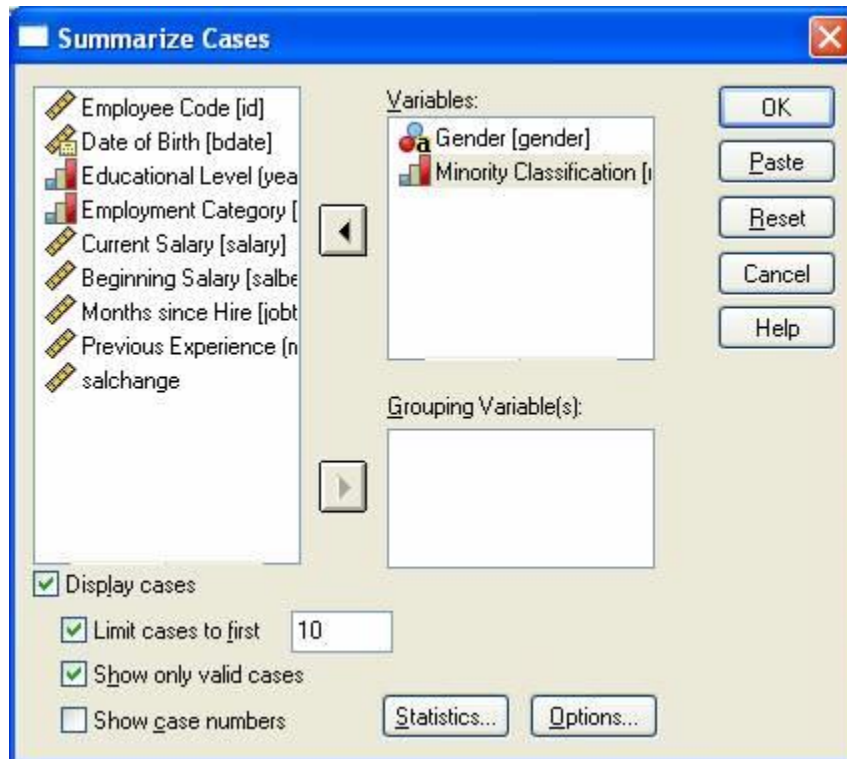


Clicking on the **If** button opens the *Select Cases: If* dialog box. Here, we select all of the cases in the dataset that meet a specific criterion: employees that have worked at the company for greater than six years (72 months) are selected. After this selection has been made, subsequent analyses will use only this subset of the data. Because we selected the *Filter* option in the previous dialog box, SPSS will indicate the inactive cases in the Data Editor by placing a slash over the row number. To select the entire dataset again, return to the *Select Cases* dialog box and select the *All Cases* option.

3.7 Listing Cases

You may sometimes want to print a list of your cases and the values of variables associated with each case, or perhaps a list of only some of the cases and variables. For example, if you want to visually examine the gender and minority status of each person in your dataset, you can generate a list of only these variables in the Output Viewer. This can be done by using the *Summarize Cases* menu option, available under the *Analyze* menu item:

Analyze
Reports
Case Summaries . . .



The above example requests a listing of each person's gender and minority status. The option *Limit cases to first* has been checked, which allows you to request a listing of only the first 10 cases. If you sorted the dataset earlier in this tutorial, then the Output Viewer would show this table:

Case Summaries^a

	Gender	Minority Classification
1	Female	No
2	Female	No
3	Female	No
4	Female	No
5	Female	No
6	Female	Yes
7	Female	No
8	Female	No
9	Female	No
10	Female	Yes
Total	N	10
		10

a. Limited to first 10 cases.

Conclusion

We hope you found this introductory SPSS tutorial enjoyable and useful! To explore further topics in SPSS, you may wish to consult the following resources:

Our SPSS FAQs (<https://stat.utexas.edu/software-faqs/spss>): Scroll down to the bottom of the FAQ list to find over fifty answers to SPSS questions we've answered for the University of Texas Community.

Our SPSS tutorials (<https://stat.utexas.edu/training/software-tutorials>): This tutorial is just one of the tutorials we offer on SPSS. Click the above link to find our other SPSS tutorials.

The SPSS website (www.spss.com): The SPSS site contains a variety of useful information, including the SPSS Answer Net (which you can search for technical FAQs) and listings of SPSS macros and algorithms.

Garson's Statnotes (www2.chass.ncsu.edu/garson/pa765/statnote.htm): Includes in-depth discussions of a variety of introductory, intermediate, and advanced statistical topics; most are illustrated using SPSS, including annotated SPSS output.

And, finally, members of UT-Austin can schedule an appointment with a statistical consultant or send e-mail to stat.consulting@austin.utexas.edu . See stat.utexas.edu/consulting/ for more details about our consulting services.