

SPSS for Windows: Getting Started

- [Section 1: Overview](#)
 - [Introduction to SPSS](#)
 - [Overview of SPSS for Windows](#)
 - [Section 2: Entering Data in SPSS](#)
 - [Starting SPSS](#)
 - [The Data Editor](#)
 - [The Syntax Editor](#)
 - [The Output Viewer](#)
 - [Importing Data from Excel Files](#)
 - [Importing Data from ASCII Files](#)
 - [Section 3: Modifying and Organizing Data in SPSS](#)
 - [Creating and Defining Data](#)
 - [Inserting Cases and Variables](#)
 - [Computing New Variables](#)
 - [Recoding Variables](#)
 - [Sorting Cases](#)
 - [Selecting Cases](#)
 - [Listing Cases](#)
-

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. Throughout these modules, a single dataset, *Employee data.sav*, is used for all examples. This example dataset that 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 10.0, 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/rc/software/> to get more information about obtaining the latest version of SPSS.

This document also contains some information about using keystrokes (or "accelerator keys") in SPSS 10. For more information, see our [SPSS 10 for Windows Keystroke Manual](#).

Section 1: Overview

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 merging and aggregating datasets. SPSS also has a number of ways to summarize and display data in the form of tables and graphs.

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

Starting SPSS

To start SPSS, go to the *Start* icon under Windows 95, Windows 98, Windows 2000, and Windows NT. You should find an SPSS icon under the *Programs* menu item. You can also start SPSS by double-clicking on an SPSS file.

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 and contains the data. You can access the second sheet by clicking on the tab labeled *Variable View* and 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 dataset that is stored with the dataset. Unlike most spreadsheets, the Data Editor can only have one dataset

open at a time. However, beginning with version 10.0, you can open multiple Data Editors at one time, each of which contains a separate dataset. Datasets that are currently open are called *working datasets* and all data manipulations, statistical functions, and other SPSS procedures operate on these datasets. The Data Editor contains several menu items that are useful for performing various operations on your data. Here is the Data Editor, containing the *Employee data.sav* dataset:

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority
1	1	m	02/03/52	15	3	\$57,000	\$27,000	98	144	0
2	2	m	05/23/58	16	1	\$40,200	\$18,750	98	36	0
3	3	f	07/26/29	12	1	\$21,450	\$12,000	98	381	0
4	4	f	04/15/47	8	1	\$21,900	\$13,200	98	190	0
5	5	m	02/09/55	15	1	\$45,000	\$21,000	98	138	0
6	6	m	08/22/58	15	1	\$32,100	\$13,500	98	67	0
7	7	m	04/26/56	15	1	\$36,000	\$18,750	98	114	0

[D](#)

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 (or type Alt+F+O+A):

File
Open...

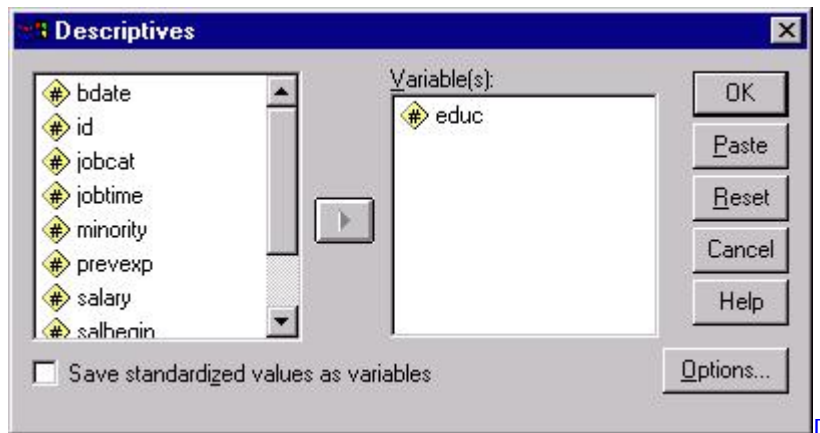
If the file you want to open is not an SPSS data file, you can often use the *Open* menu item to import that file directly into the Data Editor. If a data file is not in a format that SPSS recognizes, then try using the software package in which the file was originally created to translate it into a format that can be imported into SPSS (e.g., tab-delimited data).

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 ever writing syntax. These SPSS for Windows tutorials focus on the use of the dialog boxes to execute procedures; however, there are a couple of important 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, you should be aware of the Syntax Editor so that you can save procedures as syntax to be rerun at a latter 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. Here, only the **Paste** button in the dialog box is relevant. The process used for generating descriptive statistics is described later.



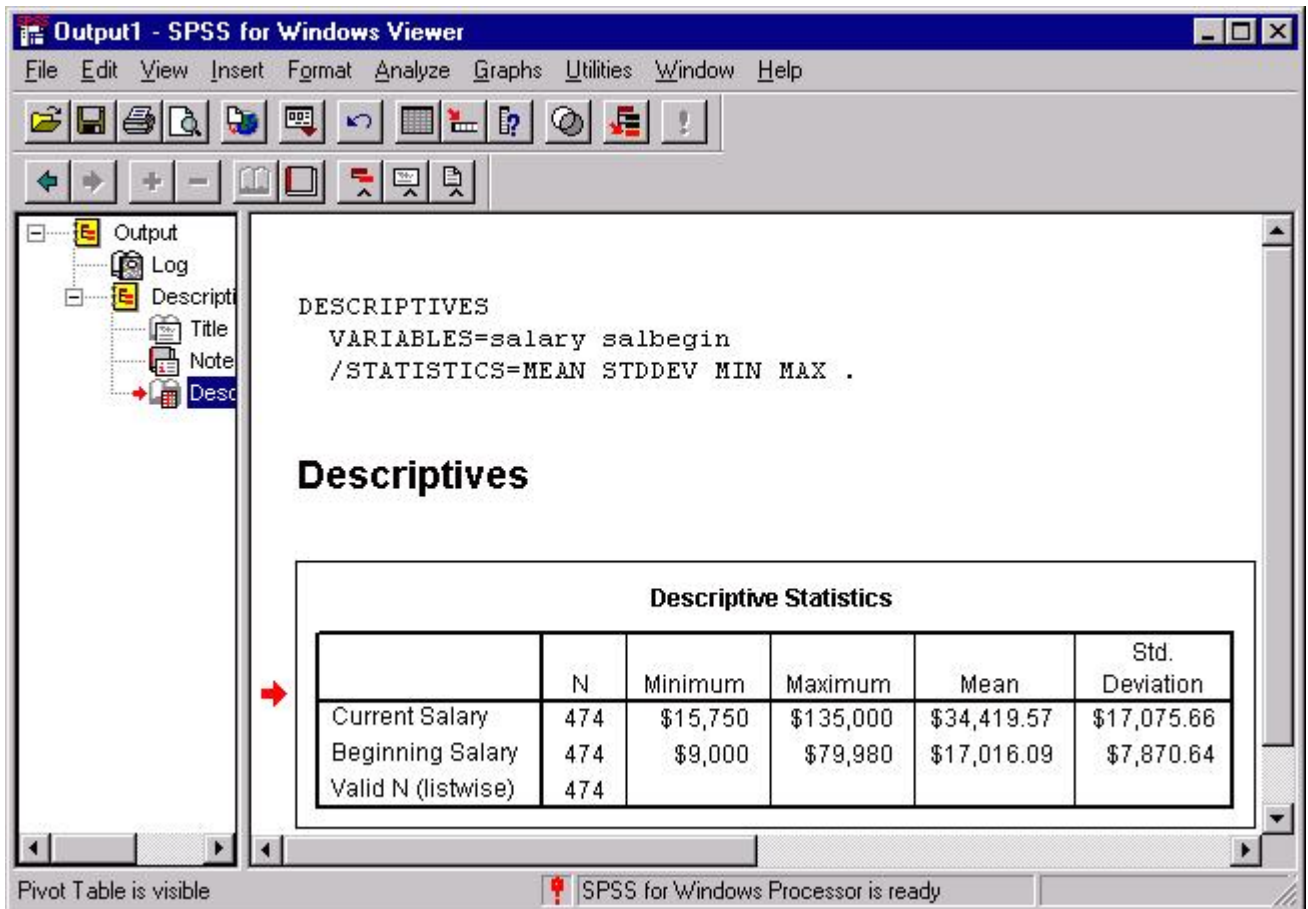
By clicking on the **Paste** button, the procedure that the above dialog box is prepared to run will be written in the form of SPSS syntax to the Syntax Editor. Thus, clicking the **Paste** (or Alt + P) button in the above example would produce the following syntax:

```
DESCRIPTIVES  
VARIABLES=educ  
/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 continuing 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.

The Output Viewer

All output from statistical analyses is printed to the Output Viewer window as well as other useful information. 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. Some other output that you may want to have printed to the Output Viewer are command syntax, titles, and error messages. The Output Viewer is shown below:



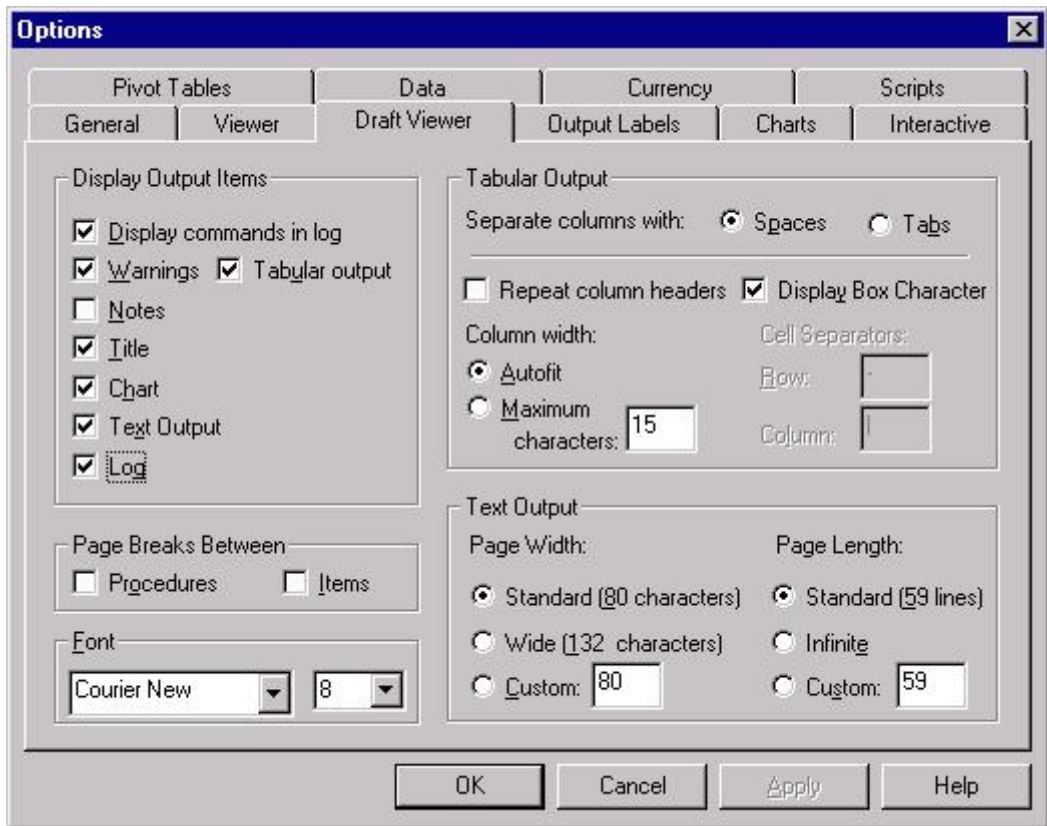
[D](#)

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 while 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 (or Alt+E+N) 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.

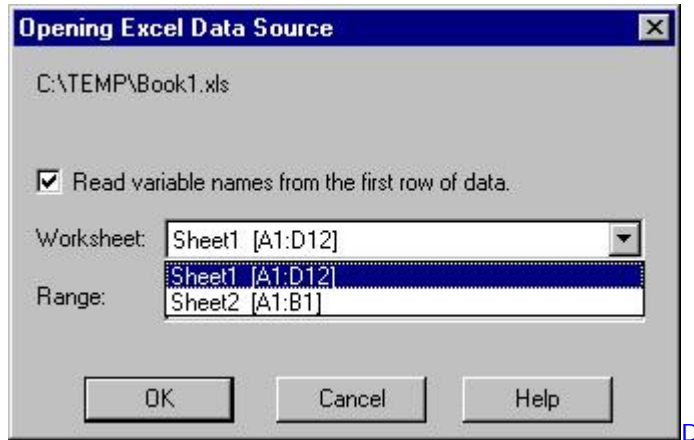
Importing Data From Excel Files

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 the fourth module of this tutorial series, [Data Manipulation and Advanced Topics](#), in the [Database Capture](#) section.

To open an Excel file, select the following menu options from the menu in the Data Editor window in SPSS (or Alt+F+O+A):

File
Open...

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 saved 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:



This dialog box allows you to select a spreadsheet from within the Excel Workbook. The drop-down menu in the example shown above offers two sheets from which to choose. As SPSS only operates on one spreadsheet at a time, you can only select one sheet from this menu. This box also gives you the option of reading variable names from the Excel Workbook directly into SPSS. Click on the *Read variable names* box to read in the first row of your spreadsheet as the variable names. If the first row of your spreadsheet does indeed contain the names of your variables and you want to import them into SPSS, these variables names should conform to SPSS variable naming conventions (eight characters or fewer, not beginning with any special characters). If you do not have variable names, use the procedure described below in [Creating and Defining Data](#) to add variable names to your dataset after you have imported your data into SPSS. 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.

If you are using a version of SPSS that was released prior to SPSS 10.0, there are a few additional steps that are necessary for opening an Excel spreadsheet directly into SPSS. You will need to save the file as an Excel version 4.0 or lower *Excel Worksheet* which is a file containing a single spreadsheet. More recent versions of Excel use the *Excel Workbook* format, which contains several spreadsheets. Because SPSS only allows one dataset to be active at any given time, it can read Excel spreadsheets which are a single spreadsheet but not Excel Workbooks which are several spreadsheets. To save an Excel file as a Excel Worksheet, choose the following from the menus (or Alt+F+A):

File
Save As...

After you assign the file a new name in the *File name* box and choose a location on disk with the *Save in* box, make sure you select the *Microsoft Excel 4.0 Worksheet (*.xls)* option from the *Save as type* pull-down menu. You will receive the following warning: "The selected file type does not support workbooks that contain multiple sheets." This warning is letting you

know that only the visible worksheet in the Excel Workbook will be saved --not all of the sheets in the workbook. Click the **OK** button here if the spreadsheet you want to import into SPSS is currently the visible sheet in your Workbook. After saving the file in this format, be sure to close the file in Excel because SPSS cannot open a file that is currently open in Excel. At this point, the file is ready to be opened in SPSS and can be opened using the procedures for SPSS 10.0 described above with the exception that you will not be offered the option to choose from available sheets as there is only a single sheet in the Worksheet.

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. SPSS 9.0 has a *Text Import Wizard* that will help you import data in an ASCII file format. The Text Import Wizard will open automatically when a ASCII file (a file with a *.txt* or *.dat* extension) is opened using the *Open* option in the *File* menu. If the data file you want to open does not have a *.txt* or *.dat* extension but you know that it is an ASCII file, then you can open the data file by opening the Data Import Wizard from the *File* menu (or Alt+F+R):

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...

(or Alt+F+S or Alt+F+A)

Section 3: Creating and Modifying Data in SPSS

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 often confused terms: *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 job classification status. Each individual in the dataset would be assigned a value representing their job classification. For instance, we could assign custodians the value 1, clerks 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 you have a variable representing employment categories that is coded as either 1, 2, or 3 for various employment categories, say, clerical custodial and managerial, it may be unwieldy to read the output if you are constantly trying to remember which number represents the 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 Information* tab. Doing so will bring the *Variable Information* sheet to the foreground. You can also access this sheet by double-clicking one of the gray boxes 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. Finally, you can also use the keystroke Ctrl+T to toggle between the windows. Regardless of the method you use, you will see a spreadsheet organized as the one below:

The screenshot shows the SPSS Data Editor window with the 'Variable View' tab selected. The table below represents the data structure shown in the window.

	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	8	0	Date of Birth	None	None	8	Right	Scale
4	educ	Numeric	2	0	Educational Le	None	0	8	Right	Ordinal
5	jobcat	Numeric	1	0	Employment C	{1, Clerical}...	0	8	Right	Ordinal
6	salary	Dollar	8	0	Current Salary	None	\$0	8	Right	Scale
7	salbegin	Dollar	8	0	Beginning Sala	None	\$0	8	Right	Scale
8	jobtime	Numeric	2	0	Months since	None	0	8	Right	Scale
9	prevexp	Numeric	6	0	Previous Exper	None	None	8	Right	Scale
10	minority	Numeric	1	0	Minority Classi	{0, No}...	9	8	Right	Ordinal

At the bottom of the window, the status bar indicates 'SPSS Processor is ready'.

D

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:

The screenshot shows the 'Variable Type' dialog box. It contains several radio button options for variable types, with 'Numeric' selected. There are also input fields for 'Width' and 'Decimal Places', and buttons for 'OK', 'Cancel', and 'Help'.

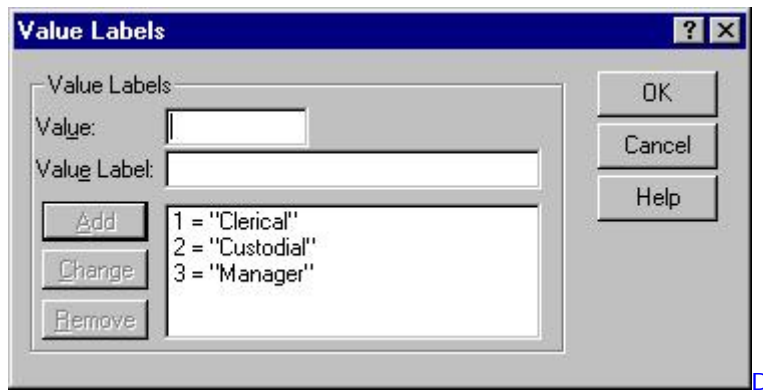
Option	Width	Decimal Places
<input checked="" type="radio"/> Numeric	1	0
<input type="radio"/> Comma		
<input type="radio"/> Dot		
<input type="radio"/> Scientific notation		
<input type="radio"/> Date		
<input type="radio"/> Dollar		
<input type="radio"/> Custom currency		
<input type="radio"/> String		

D

This box allows you to define the type of data for variables. For example, you will be presented with *Numeric*, *String*, and *Date* options among others. Thus, if you want to define the variable *jobcat*, a variable representing employment category as a string variable rather than the default variable type, numeric, you would click on the the cell in the *jobcat* row and the *Type* column, then click the gray box to produce the *Variable Type* dialog box. Here, you would choose the *String* option.

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 a 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.

Inserting and Deleting Cases and Variables

You may want to add new variables or cases to an existing dataset. The Data Editor provides menu options that allow you to do that. For example, you may want to add data about participants' ages to an existing dataset. To insert a new variable, click on the variable name to select the column in which the variable is to be inserted. To insert a case, select the row in which the case is to be added by clicking on the row's number. Clicking on either the row's number or the column's name will result in that row or column being highlighted. Next, use the insert options available in the *Data* menu in the Data Editor:

Data

Insert Variable

Insert Case

(or Alt+D+V for Variables and Alt+D+I for Cases)

If a row has been selected, choose *Insert Case* from the *Data* menu; if a column has been selected, choose, *Insert Variable*. This will produce an empty row or column in the highlighted area of the Data Editor. Existing cases and variables will be shifted down or to the right, respectively.

You may want to delete cases or variables from a dataset. To do that, select a row or column by highlighting as described above. Next, use the **Delete** key to delete the highlighted area. Or you can use the *Delete* option in the *Edit* menu to do it.

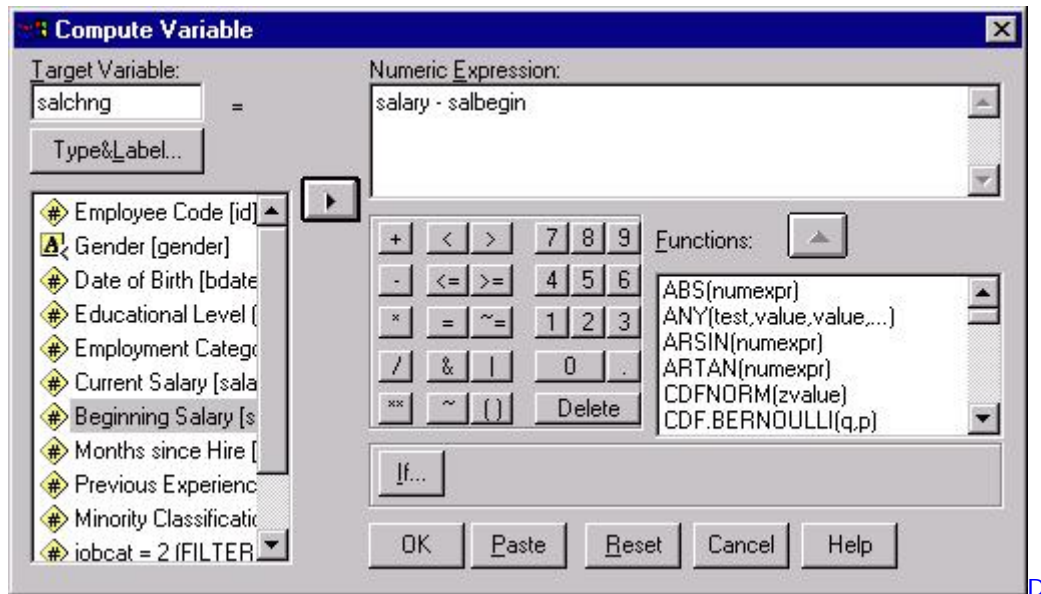
Computing New Variables

You may want to modify the values of the variables in your datasets. For example, if a dataset contained employees' salaries in terms of their beginning and current salaries but you wanted the difference between starting salary and present salary, a new variable could be computed by subtracting the starting salary from the present salary. In other situations, you may also want to transform an existing variable. For example, if data were entered as months of experience and you wanted to analyze data in terms of years on the job, then you could recompute that variable to represent experience on the job in numbers of years by dividing number of months on the job by 12.

Both variables that are created as a numeric expression of existing variables and variables whose values are modified by an operation can be computed using the *Compute* option available from the menu in the Data Editor (or Alt+T+C):

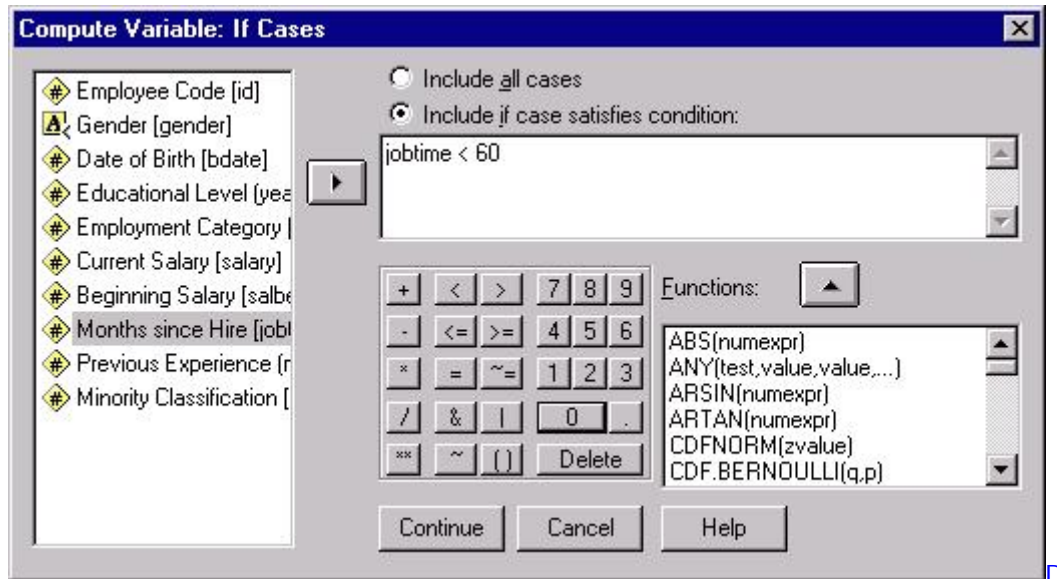
Transform Compute...

This will result in the following dialog box:



To create a new variable, type its name in the box labeled *Target Variable*. Alternatively, you may want to modify the value of an existing variable, in which case you would type its name in the box labeled *Target Variable*. In both cases, 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, *salchn*, 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 five years, you could create a condition that would compute a new variable only if an employee had begun employment within the last five years. 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 gray areas of the dialog box. Then, specify the condition for computing a new variable in the input box at the top right of this dialog box. You can either type in the condition or click on variables in the variable 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 selecting 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 five years (60 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. Thus, only cases which have fewer than sixty months, or five years, since they were hired will be included. Click the **Continue** button to return to the previous dialog box.

Recoding Variables

You can also modify the values of existing variables in your dataset. For example, if a dataset contains a variable that classifies an employee's status in three categories, but for a particular analysis you want to combine two of these classifications into a single category, then two of the values would need to be recoded into a single value so that there are two total groups.

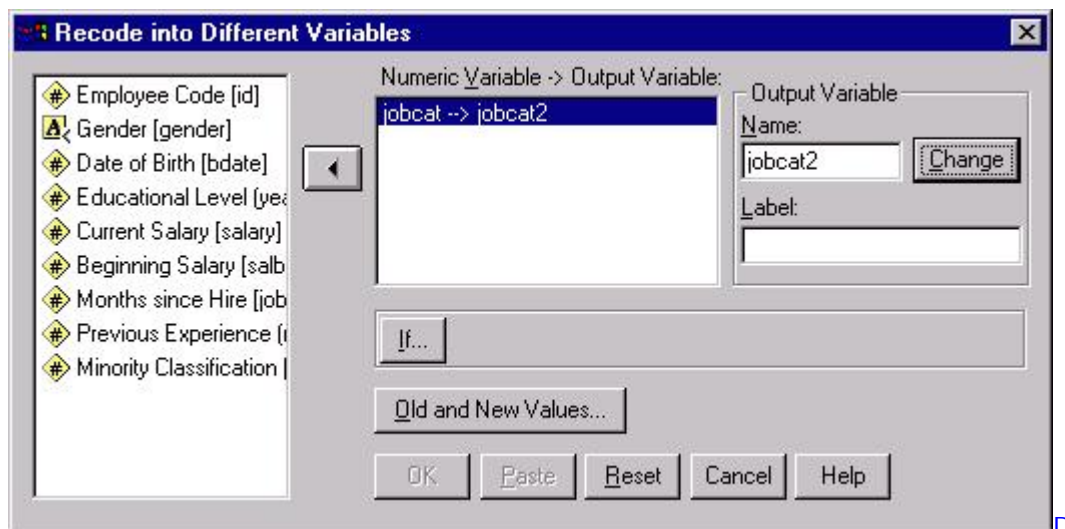
The *Recode* option (or Alt+T+R) is available from the menu in the Data Editor:

Transform

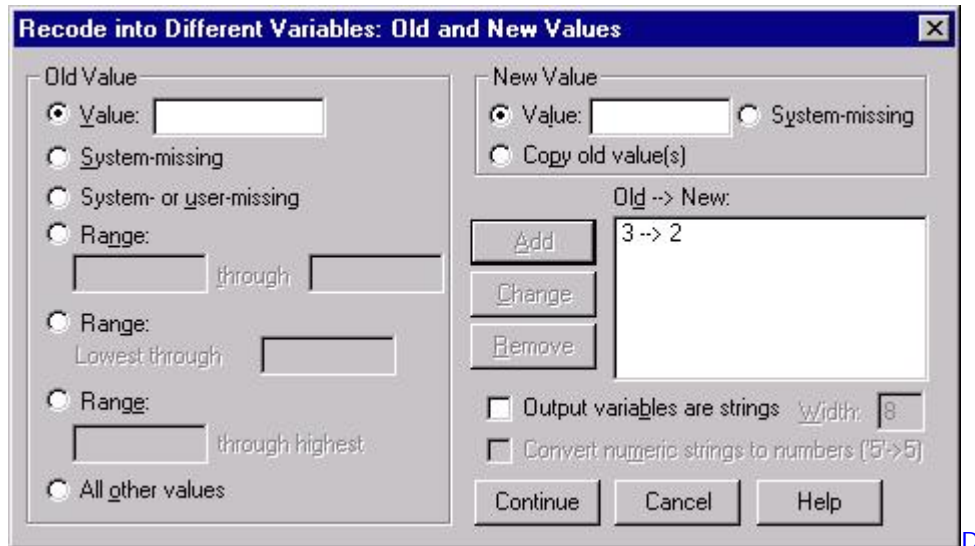
Recode

Additionally, there are two options for recoding variables in the *Recode* submenu. The *Into Same Variables* (Alt+T+R+S) option changes the values of the existing variables, whereas the *Into Different Variables* (Alt+T+R+D) 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. You should use the *Into Different Variables* option, because you may change your mind about your recoding scheme at a later date. Thus, if you do change your mind, you still have the original values.

The following example illustrates the use of the *Recode* option to recode values into a new variable. When that option is selected from the menu, the following dialog box will appear:



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. After a new variable name has been supplied, click on the button labeled *Old and new Values*. This will result in the following dialog box:



The above dialog box is the same regardless of whether you are recoding values into the same variable or creating a new variable. 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.

Continuing with the above example, a variable with three values, such as *jobcat*, could be recoded into a variable with two values by recoding one of the 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. For example, by entering 3 in the box labeled *Old Value* and entering 2 in the box labeled *New Value* then clicking **Add**, all of the cases labeled 3 would take on the value 2. This can be repeated for as many of the values as necessary.

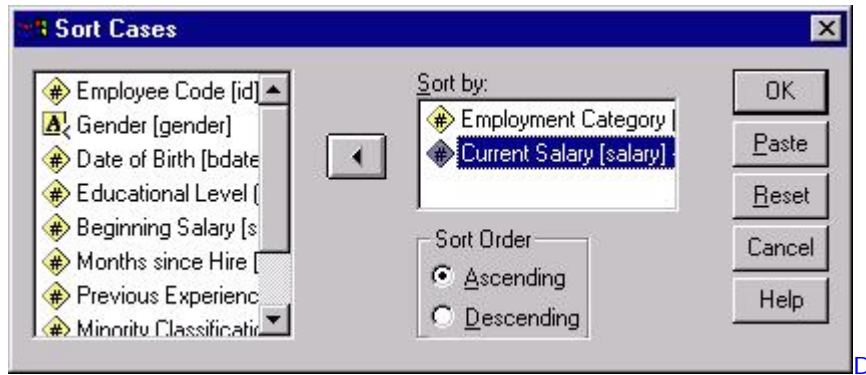
Values can also be recoded conditionally. The process for recoding values on the basis of a condition is essentially identical to the process for conditionally computing new variables discussed in the previous section: when you click on the **If** button in the main *Recode* dialog box, the same dialog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

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* (or Alt+ D+O) 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. For example, if *jobcat* was the first variable entered, followed by *salary*, the data would first be sorted by *jobcat*, then, within each of the job categories, data would be sorted by *salary*.

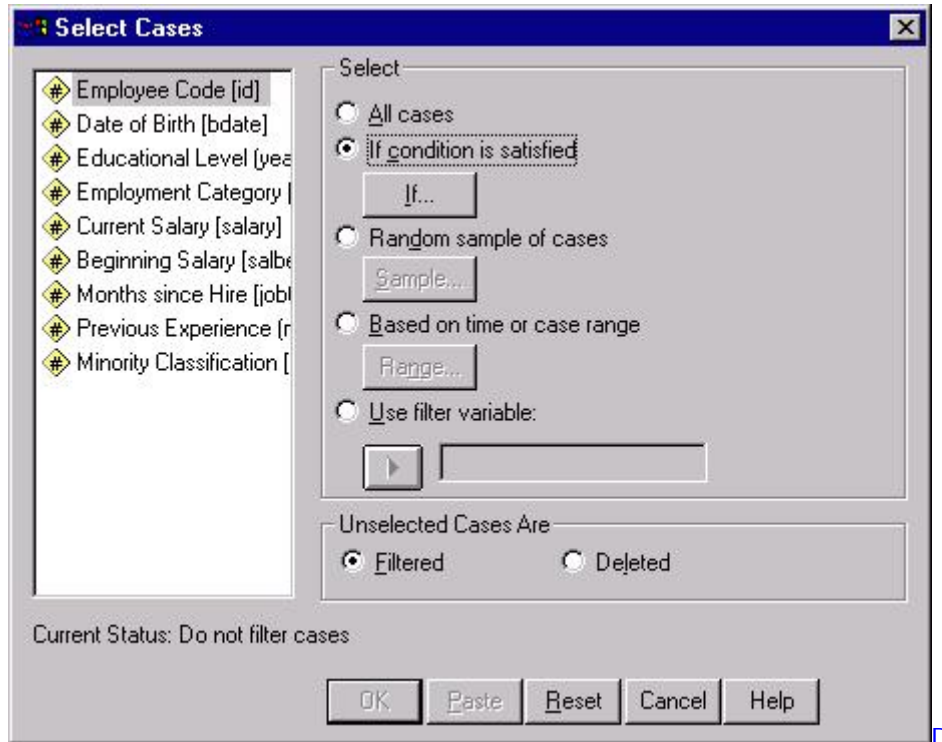
Selecting Cases

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 didn't want from the dataset. The *Select Cases* option (or Alt+D+C) 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.



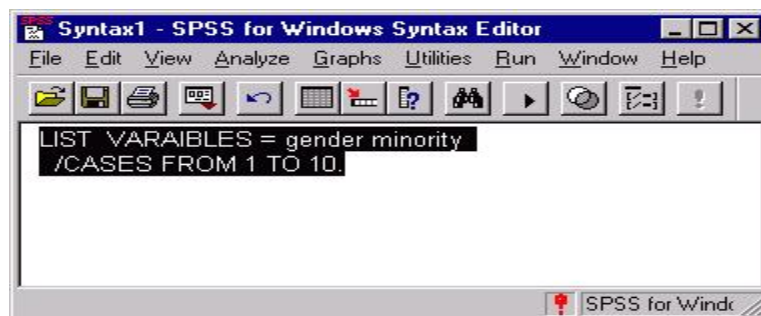
Selecting one of these options will produce a second dialog box that prompts you for the particular specifications in which you are interested. For example, selecting the *If condition is satisfied* option and clicking on the **If** button (as was done in the example) results in a second dialog box, as shown below. The portion of the dialog box labeled *Unselected Cases Are* gives you the option of temporarily or permanently removing data from the dataset. The *Filtered* 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. If the *Deleted* option is selected, the unselected cases will be removed from the working dataset. If the dataset is subsequently saved, these cases will be permanently deleted.



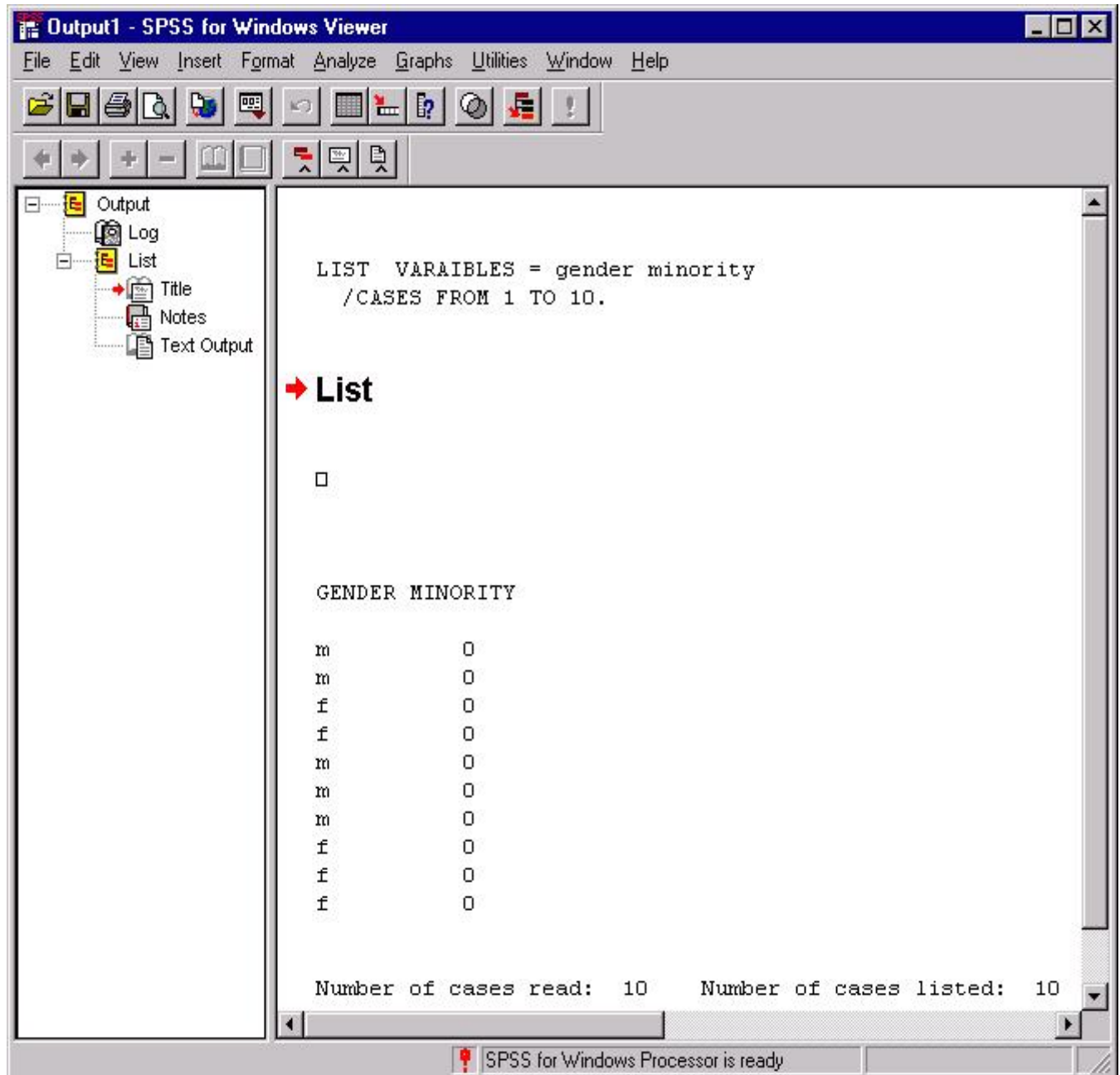
The above example selects all of the cases in the dataset that meet a specific criterion: employees that have worked at the company for greater than six years (72 months) will be selected. After this selection has been made, subsequent analyses will use only this subset of the data. If you have chosen 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.

Listing Cases

You may sometime 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 have two variables that you want to examine visually, but this cannot be done because they are at very different places in your dataset, you could generate a list of only these variables in the Output Viewer. The procedure for doing this cannot be performed using dialog boxes and is available only through command syntax. The syntax for generating a list of cases is shown in the Syntax Editor window below. The variable names shown in lower case below instruct SPSS which variables to list in the output. Or, you can type in the command **ALL** in place of variables names, which will produce a listing of all of the variables in the file. The subcommand **/CASES FROM 1 TO 10**, is an instruction to SPSS to print only the first ten cases. If this instruction were omitted, all cases would be listed in the output.



To execute this command, first highlight the selection by pressing on your mouse button while dragging the arrow across the command or commands that you want to execute. Next, click on the icon with the black, right-facing arrow on it. Or, you can choose a selection from the *Run* menu. Executing the command will print the list of variables, *gender* and *minority* in the above example, to the Output Viewer. The *Output Viewer* is the third window with which you should be familiar. It is the window in which all output will be printed. The Output Viewer is shown below, containing the text that would be generated from the above syntax.



SPSS for Windows: Descriptive and Inferential Statistics

- [Section 4: Summarizing Data](#)
 - [Descriptive Statistics](#)
 - [Frequencies](#)
 - [Crosstabulation](#)
- [Section 5: Inferential Statistics](#)
 - [Chi-Square](#)
 - [T test](#)
 - [Correlation](#)
 - [Regression](#)
 - [General Linear Model](#)

This document is the second module of a four module tutorial series. This document describes the use of SPSS to obtain descriptive and inferential statistics. In this module, you will be introduced to procedures used to obtain several descriptive statistics, frequency tables, and crosstabulations in the first section. In the second section, the Chi-square test of independence, independent and paired sample *t* tests, bivariate and partial correlations, regression, and the general linear model will be covered. If you are not familiar with SPSS or need more information about how to get SPSS to read your data, consult the first module of this four part tutorial, [SPSS for Windows: Getting Started](#). This set of documents uses a sample dataset, *Employee data.sav*, that SPSS provides. It can be found in the root SPSS directory. If you installed SPSS in the default location, then this file will be located in the following location: C:\Program Files\SPSS\Employee Data.sav.

Some users prefer to use keystrokes to navigate through SPSS. Information on common keystrokes are available in our [SPSS 10 for Windows Keystroke Manual](#).

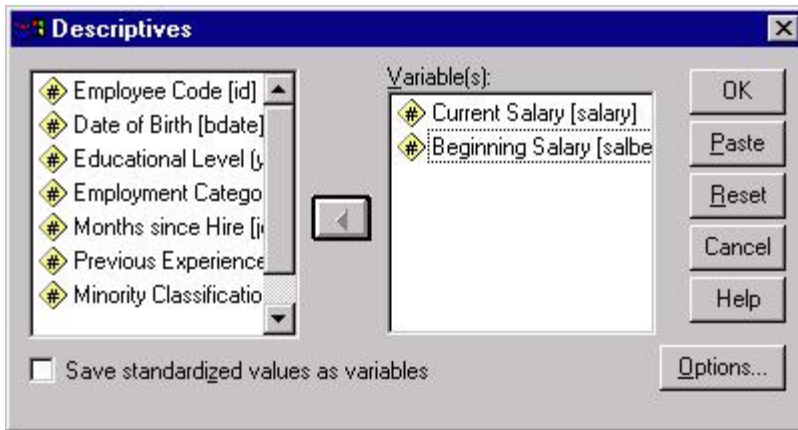
Section 4: Summarizing Data

Descriptive Statistics

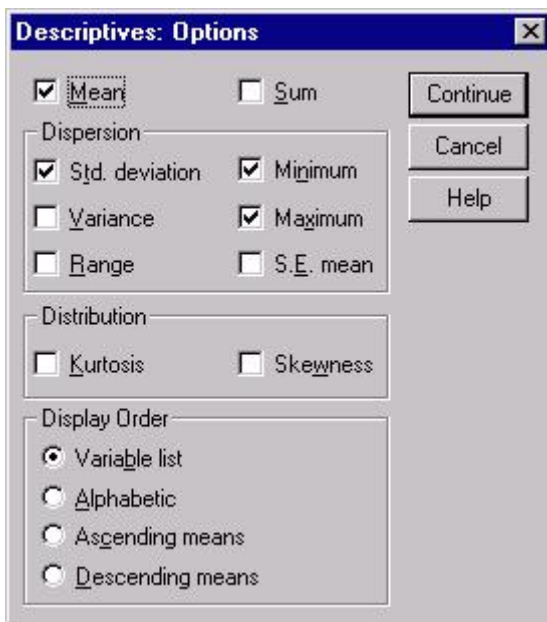
A common first step in data analysis is to summarize information about variables in your dataset, such as the averages and variances of variables. Several summary or descriptive statistics are available under the *Descriptives* option available from the *Analyze* and *Descriptive Statistics* menus:

```
Analyze
  Descriptive Statistics
    Descriptives...
```

After selecting the *Descriptives* option, the following dialog box will appear:



This dialog box allows you to select the variables for which descriptive statistics are desired. To select variables, first click on a variable name in the box on the left side of the dialog box, then click on the arrow button that will move those variables to the *Variable(s)* box. For example, the variables *salbegin* and *salary* have been selected in this manner in the above example. To view the available descriptive statistics, click on the button labeled **Options**. This will produce the following dialog box:



Clicking on the boxes next to the statistics' names will result in these statistics being displayed in the output for this procedure. In the above example, only the default statistics have been selected (mean, standard deviation, minimum, and maximum), however, there are several others that could be selected. After selecting all of the statistics you desire, output can be generated by first clicking on the **Continue** button in the *Options* dialog box, then clicking on the **OK** button in the *Descriptives* dialog box. The statistics that you selected will be printed in the Output Viewer. For example, the selections from the preceding example would produce the following output:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Current Salary	174	\$15,750	\$135,000	\$34,419.57	\$17,016.09
Beginning Salary	174	\$9,000	\$79,980	\$17,016.09	\$7,870.64
Valid N (listwise)	174				

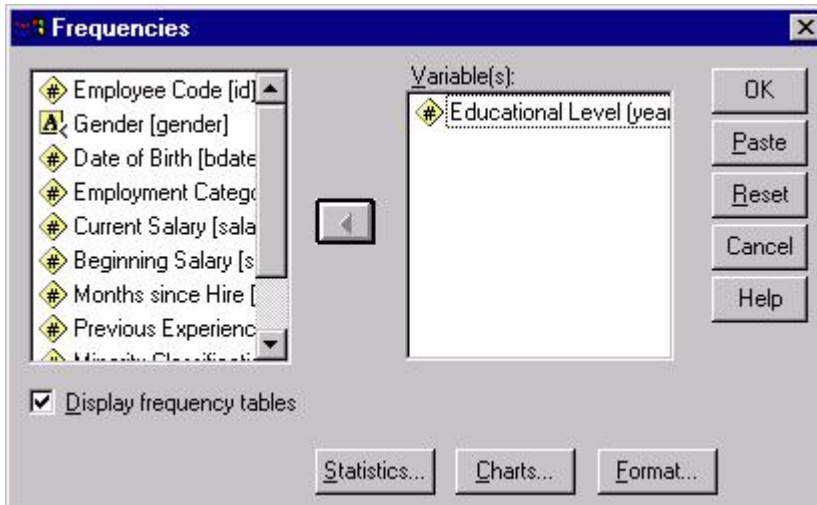
This output contains several pieces of information that can be useful to you in understanding the descriptive qualities of your data. The number of cases in the dataset is recorded under the column labeled *N*. Information about the range of variables is contained in the *Minimum* and *Maximum* columns. For example, beginning salaries ranged from \$9000 to \$79,980 whereas current salaries range from \$15,750 to \$135,000. The average salary is contained in the *Mean* column. Variability can be assessed by examining the values in the *Std.* column. The standard deviation measures the amount of variability in the distribution of a variable. Thus, the more that the individual data points differ from each other, the larger the standard deviation will be. Conversely, if there is a great deal of similarity between data points, the standard deviation will be quite small. The standard deviation describes the standard amount variables differ from the mean. For example, a starting salary with the value of \$24,886.73 is one standard deviation above the mean in the above example in which the variable, *salary* has a mean of \$17,016.09 and a standard deviation of \$7,870.64. Examining differences in variability could be useful for anticipating further analyses: in the above example, it is clear that there is much greater variability in the current salaries than beginning salaries. Because equal variances is an assumption of many inferential statistics, this information is important to a data analyst.

Frequencies

While the descriptive statistics procedure described above is useful for summarizing data with an underlying continuous distribution, the *Descriptives* procedure will not prove helpful for interpreting categorical data. Instead, it is more useful to investigate the numbers of cases that fall into various categories. The *Frequencies* option allows you to obtain the number of people within each education level in the dataset. The *Frequencies* procedure is found under the *Analyze* menu:

Analyze
 Descriptives Statistics
 Frequencies...

Selecting this menu item produces the following dialog box:

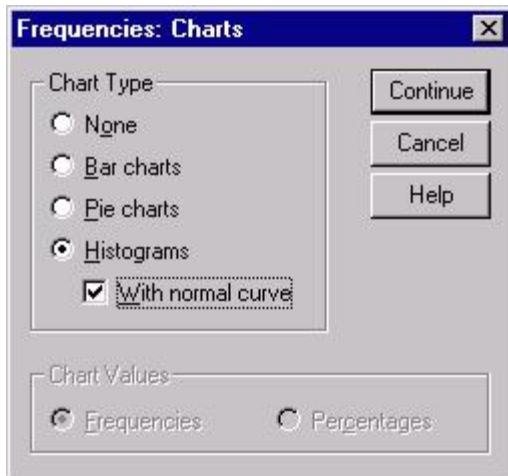


Select variables by clicking on them in the left box, then clicking the arrow in between the two boxes. Frequencies will be obtained for all of the variables in the box labeled *Variable(s)*. This is the only step necessary for obtaining frequency tables; however, there are several other descriptive statistics available, many of which are described in the preceding section. The example in the above dialog box would produce the following output:

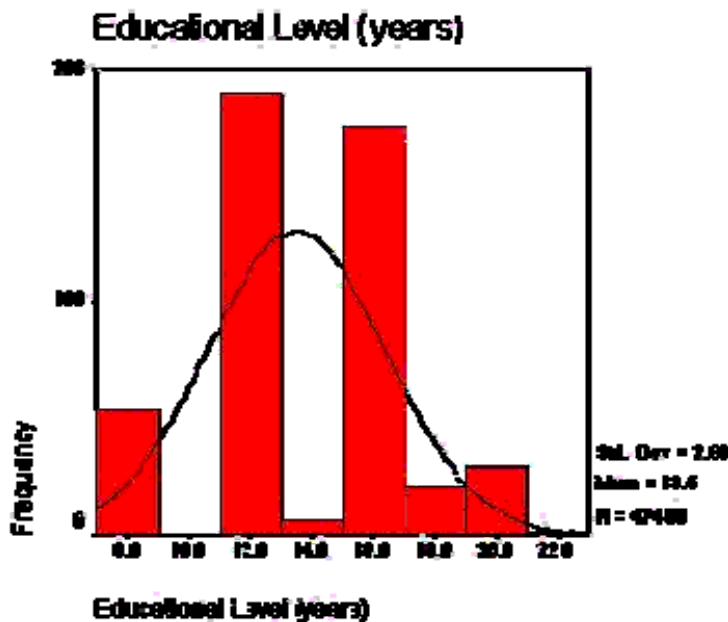
Educational Level (years)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8	63	11.2	11.2	11.2
	12	190	40.1	40.1	51.3
	14	8	1.3	1.3	52.6
	16	118	24.6	24.6	77.0
	18	68	12.4	12.4	89.5
	17	11	2.3	2.3	91.8
	19	9	1.9	1.9	93.7
	19	27	6.7	6.7	99.4
	20	2	.4	.4	99.8
	21	1	.2	.2	100.0
	Total	474	100.0	100.0	

Clicking on the **Statistics** button produces a dialog box with several additional descriptive statistics. Clicking on the **Charts** button produces the following box which allows you to graphically examine their data in several different formats:



Each of the available options provides a visual display of the data. For example, clicking on the *Histograms* button with its suboption, *With normal curve*, will provide you with a chart similar to that shown below. This will allow you to assess whether your data are normally distributed, which is an assumption of several inferential statistics. You can also use the *Explore* procedure, available from the *Descriptives* menu, to obtain the *Kolmogorov-Smirnov test*, which is a hypothesis test to determine if your data are normally distributed.



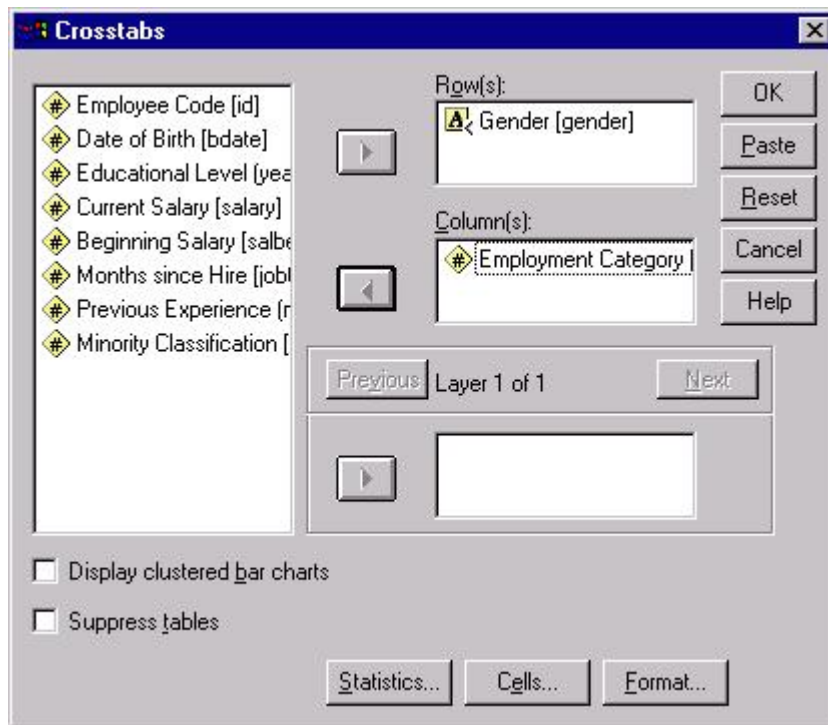
Crosstabulation

While frequencies show the numbers of cases in each level of a categorical variable, they do not give information about the relationship between categorical variables. For example, frequencies can give you the number of men and women in a company AND the number of people in each employment category, but not the number of men and women IN each employment category. The *Crosstabs* procedure is useful for investigating this type of

information because it can provide information about the intersection of two variables. The number of men and women in each of three employment categories is one example of information that can be crosstabulated. The *Crosstabs* procedure is found in the *Analyze* menu in the Data Editor window:

Analyze
Descriptive Statistics
Crosstabs...

After selecting *Crosstabs* from the menu, the dialog box shown above will appear on your monitor. The box on the left side of the dialog box contains a list of all of the variables in the working dataset. Variables from this list can be selected for rows, columns, or layers in a crosstabulation. For example, selecting the variable *gender* for the rows of the table and *jobcat* for the columns would produce a crosstabulation of gender by job category.



The options available by selecting the **Statistics** and **Cells** buttons provide you with several additional output features. Selecting the **Cells** button will produce a menu that allows you to add additional values to your table. For example, the dialog box shown below illustrates an example in which *Expected* option in the *Counts* box and the *Row*, *Column*, and *Total* options in the *Percentages* box have been selected.

Crosstabs: Cell Display

Counts

☒ Observed

☒ Expected

Percentages

☒ Row

☒ Column

☒ Total

Residuals

☐ Unstandardized

☐ Standardized

☐ Adj. standardized

Continue

Cancel

Help

The combination of the two dialog boxes shown above will produce the following output table:

Gender * Employment Category Crosstabulation

			Employment Category			Total
			Clerical	Custodial	Manager	
Gender	Female	Court	200	8	10	218
		Expected Court	185.4	12.3	38.3	218.0
		% within Gender	93.4%	.0%	4.8%	100.0%
		% within Employment Category	58.7%	.0%	17.9%	45.8%
		% of Total	43.5%	.0%	2.1%	45.8%
	Male	Court	157	27	74	258
		Expected Court	197.6	14.7	45.7	258.0
		% within Gender	60.9%	10.5%	28.7%	100.0%
		% within Employment Category	43.3%	100.0%	88.1%	54.4%
		% of Total	33.1%	6.7%	16.8%	64.4%
Total		Court	363	27	84	474
		Expected Court	363.0	27.0	84.0	474.0
		% within Gender	78.8%	5.7%	17.7%	100.0%
		% within Employment Category	100.0%	100.0%	100.0%	100.0%
		% of Total	78.8%	5.7%	17.7%	100.0%

The crosstabulation statistics provide several interesting observations about the data. In the above table, there appears to be an association between gender and employment category as the expected values, which are the values expected by chance, and the actual counts are different from each other. The following section will discuss how to further examine this relationship with inferential statistics.

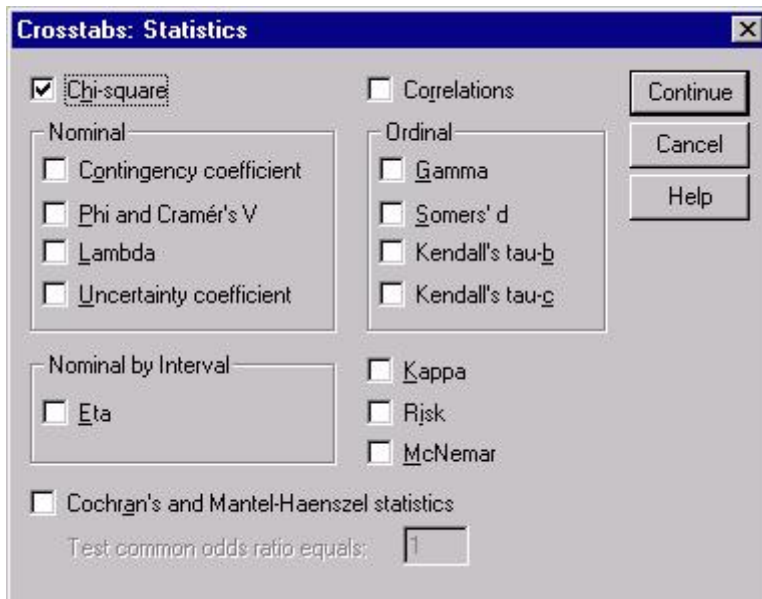
Section 5: Inferential Statistics

Chi-Square Test

The Chi-square test for independence is used in situations where you have two *categorical variables*. A categorical variable is a qualitative variable in which cases are classified in one and only one of the possible levels. A classic example is gender, in which cases are classified in one of two possible levels. The example in the above section, in which *Gender* and *Employment Category* are crosstabulated using the SPSS *Crosstabs* procedure, is an example of data with which you could conduct a *Chi-square test of independence* testing the null hypothesis that there is no relationship between the two variables.

For instance, you could conduct a test of the hypothesis that there is no relationship between *Gender* and *Employment Category*. If this hypothesis were true, you would expect that the proportion of men and women would be the same within each level of *Employment Category*. In other words, there should be little difference between *observed* and *expected* values, where the expected values represent the numbers that would be in each cell when the variables are independent of each other. The difference between observed and expected values is the basis of the Chi-square statistic: it evaluates the likelihood that the differences between the observed and expected values would occur under the null hypothesis that there is no difference between these values. The expected values can be obtained by clicking on the **Cells** box in the *Crosstabs* dialog box, as described in the preceding section. Examining the table above, it appears that it is indeed the case that gender and employment category are not independent of each other. It appears that there are more women in clerical positions than would be expected by chance, whereas there are more men in custodial and managerial positions than would be expected by chance. Conducting a Chi-square test of independence would tell us if the observed pattern is statistically different from the pattern expected due to chance.

The Chi-square test of independence can be obtained through the *Crosstabs* dialog boxes that were used above to get a crosstabulation of the data. After opening the *Crosstabs* dialog box as described in the preceding section, click the **Statistics** button to get the following dialog box:



By clicking on the box labeled *Chi-Square*, you will obtain the Chi-square test of independence for the variables you have crosstabulated. This will produce the following table in the Output Viewer:

CHI-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	79.277^a	2	.000
Likelihood Ratio	95.433	2	.000
N of Valid Cases	474		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.30.

Inspecting the table in the previous section, it appears that the two variables, gender and employment category, are related to each other in some way. This finding is implicated by the substantial differences in the observed and expected counts: these differences represent the difference between values expected if gender and employment classification were independent of each other (expected counts) and the actual numbers of cases in each cell (observed counts). For example, if gender and employment classification were unrelated, then it is expected that 38.3 women would be in the manager classification as opposed to the observed number, 10. In this example, the expected value of 38.3 represents the fact that 45.6% of the cases in this dataset are women, so it is expected that 45.6% of the 84 managers in the dataset would also be women if gender and employment classification were independent of each other. The output above provides a statistical hypothesis test for the hypothesis that gender and employment category are independent of each other. The large Chi-Square statistic (79.28) and its small significance level ($p < .000$) indicates that it is very unlikely that these variables are independent of each other. Thus, you can conclude that there is a relationship between a person's gender and their employment classification.

T tests

The t test is a useful technique for comparing mean values of two sets of numbers. The comparison will provide you with a statistic for evaluating whether the difference between two means is statistically significant. T tests can be used either to compare two independent groups (independent-samples t test) or to compare observations from two measurement occasions for the same group (paired-samples t test). To conduct a t test, your data should be a sample drawn from a continuous underlying distribution. If you are using the t test to compare two groups, the groups should be randomly drawn from normally distributed and independent populations. For example, if you were comparing clerical and managerial salaries, the *independent populations* are clerks and managers, which are two nonoverlapping groups. If you have more than two groups or more than two variables in a single group that you want to compare, you should use one of the General Linear Model procedures in SPSS, which are described below.

There are three types of t tests; the options are all located under the *Analyze* menu item:

Analyze

Compare Means

One-Sample T test...

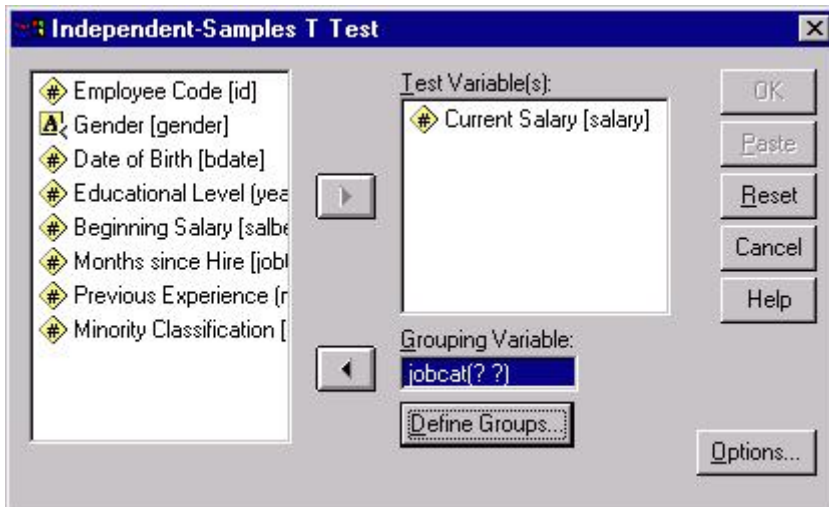
Independent-Samples T test...

Paired-Samples T test...

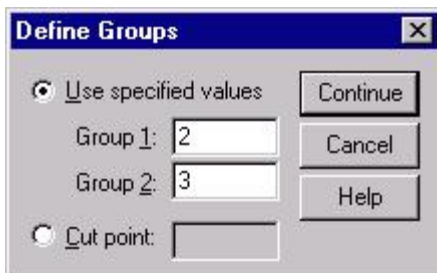
While each of these t tests compares mean values of two sets of numbers, they are designed for distinctly different situations:

- The *one-sample t test* is used to compare a single sample with a population value. For example, a test could be conducted to compare the average salary of managers within a company with a value that was known to represent the national average for managers.
- The *independent-sample t test* is used to compare two groups' scores on the same variable. For example, it could be used to compare the salaries of clerks and managers to evaluate whether there is a difference in their salaries.
- The *paired-sample t test* is used to compare the means of two variables within a single group. For example, it could be used to see if there is a statistically significant difference between starting salaries and current salaries among the custodial staff in an organization.

To conduct an independent sample t test, first select the menu option shown above, to produce the following dialog box:



To select variables for the analysis, first highlight them by clicking on them in the box on the left. Then move them into the appropriate box on the right by clicking on the arrow button in the center of the box. Your independent variable should go in the *Grouping Variable* box, which is a variable that defines which groups are being compared. For example, because employment categories are being compared in this analysis, the *jobcat* variable is selected. However, because *jobcat* has more than two levels, you will need to click on **Define Groups** to specify the two levels of *jobcat* that you want to compare. This will produce another dialog box as is shown below:



Here, the groups to be compared are limited to the groups with the values 2 and 3, which represent the clerical and managerial groups. After selecting the groups to be compared, click the **Continue** button, and then click the **OK** button in the main dialog box. The above choices will produce the following output:

Group Statistics

Employment Category		N	Mean	Std. Deviation	Std. Error Mean
Current Salary	Clerical	27	\$30,938.89	\$2,114.62	\$406.96
	Manager	64	\$63,977.80	\$18,244.78	\$1,990.67

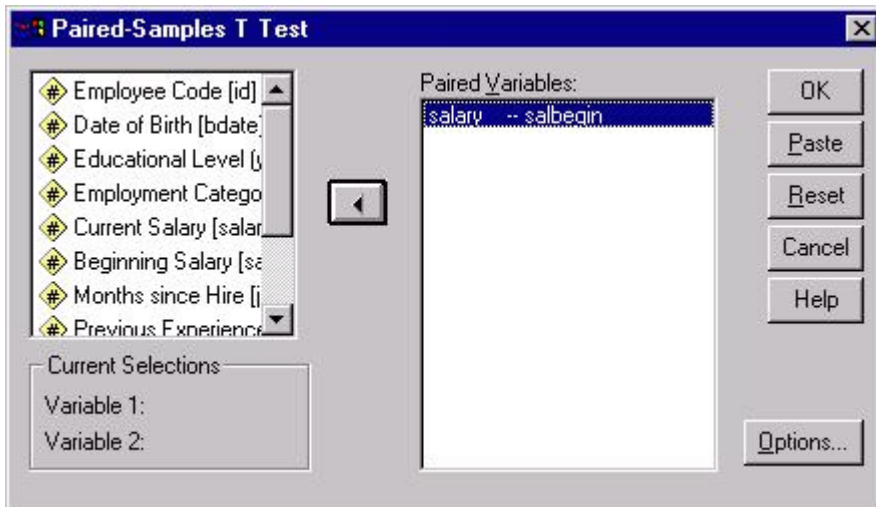
Independent Samples Test

		Levene's Test for Equality of Variances		t-Test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Current Salary	Equal variances assumed	28.2	.000	-8.38	108	.000	-\$33,038	\$3,630	-\$40,034	-\$26,044
	Equal variances not assumed			-16.3	89.6	.000	-\$33,038	\$2,032	-\$37,078	-\$29,002

The first output table, labeled *Group Statistics*, displays descriptive statistics. The second output table, labeled *Independent Samples Test*, contains the statistics that are critical to evaluating the current research question. This table contains two sets of analyses: the first assumes equal variances and the second does not. To assess whether you should use the statistics for equal or unequal variances, use the significance level associated with the value under the heading, *Levene's Test for Equality of Variances*. It tests the hypothesis that the variances of the two groups are equal. A small value in the column labeled *Sig.* indicates that this hypothesis is false and that the groups do indeed have unequal variances. In the above case, the small value in that column indicates that the variance of the two groups, clerks and managers, is not equal. Thus, you should use the statistics in the row labeled *Equal variances not assumed*.

The SPSS output reports a *t statistic* and *degrees of freedom* for all *t* test procedures. Every unique value of the *t* statistic and its associated degrees of freedom have a significance value. In the above example in which the hypothesis that clerks and managers do not differ in their salaries, the *t* statistic under the assumption of unequal variances has a value of -16.3, and the degrees of freedom has a value of 89.6 with an associated significance level of .000. The significance level tells us that the probability that there is no difference between clerical and managerial salaries is very small: specifically, less than one time in a thousand would we obtain a mean difference of \$33,038 or larger between these groups if there were really no differences in their salaries.

To obtain a paired-samples *t* test, select the menu items described above and the following dialog box will appear:



The above example illustrates a t test between the variables *salbegin* and *salary* which represent employees' beginning salary and their current salary. To set up a paired-samples t test as in the above example, click on the two variables that you want to compare. The variable names will appear in the section of the box labeled *Current Selections*. When these variable names appear there, click the arrow in the middle of the dialog box and they will appear in the *Paired Variables* box. Clicking the **OK** button with the above variables selected will produce output for the paired-samples t test. The following output is an example of the statistics you would obtain from the above example.

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Current Salary - Beginning Salary	\$17,403.48	\$10,814.62	\$496.73	\$16,427.41	\$18,379.98	35.04	473	.000

As with the independent samples t test, there is a t statistic and degrees of freedom that has a significance level associated with it. The t test in this example tests the hypothesis that there is no difference in clerks' beginning and current salaries. The t statistic, (35.04), and its associated significance level ($p < .000$) indicate that this is not the case. In fact, the observed mean difference of \$17,403.48 between beginning and current salaries would occur fewer than once in a thousand times if there really were no difference between clerks' beginning and current salaries.

Correlation

Correlation is one of the most common forms of data analysis both because it can provide an analysis that stands on its own, and also because it underlies many other analyses, and can be a good way to support conclusions after primary analyses have been completed.

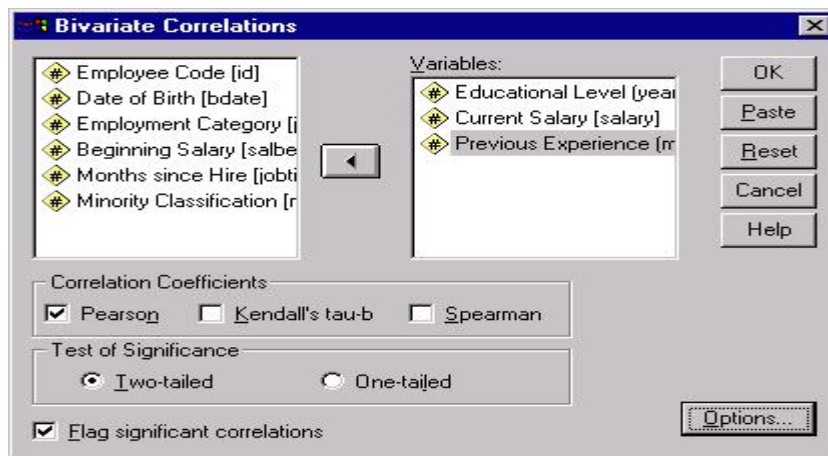
Correlations are a measure of the linear relationship between two variables. A correlation coefficient has a value ranging from -1 to 1. Values that are closer to the absolute value of 1 indicate that there is a strong relationship between the variables being correlated whereas values closer to 0 indicate that there is little or no linear relationship. The sign of a correlation coefficient describes the type of relationship between the variables being correlated. A positive correlation coefficient indicates that there is a positive linear relationship between the variables: as one variable increases in value, so does the other. An example of two variables that are likely to be positively correlated are the number of days a student attended class and test grades because, as the number of classes attended increases in value, so do test grades. A negative value indicates a negative linear relationship between variables: as one variable increases in value, the other variable decreases in value. The number of days students miss class and their test scores are likely to be negatively correlated because as the number of days of missed classed increases, test scores typically decrease.

To obtain a correlation in SPSS, start at the *Analyze* menu. Select the *Correlate* option from this menu. By selecting this menu item, you will see that there are three options for correlating variables: (1) *Bivariate*, (2) *Partial*, and (3) *Distances*. This document will cover the first two types of correlations. The *bivariate correlation* is for situations where you are interested only in the relationship between two variables. *Partial correlations* should be used when you are measuring the association between two variables but want to factor out the effect of one or more other variables.

To obtain a bivariate correlation, choose the following menu option:

Analyze
Correlate
Bivariate...

This will produce the following dialog box:



To obtain correlations, first click on the variable names in the variable list on the left side of the dialog box. Next, click on the arrow between the two white boxes which will move the selected variables into the *Variables* box. Each variable listed in the *Variables* box will be correlated with every other variable in the box. For example, with the above selections, we would obtain correlations between *Education Level* and *Current Salary*, between *Education Level* and *Previous Experience*, and between *Current Salary* and *Previous Experience*. We will maintain the default options shown in the above dialog box in this example. The first option to consider is the type of correlation coefficient. Pearson's is appropriate for continuous data as noted in the above example, whereas the other two correlation coefficients, Kendall's tau-b and Spearman's, are designed for ranked data. The choice between a one and two-tailed significance test in the *Test of Significance* box should be determined by whether the hypothesis you are testing is making a prediction about the direction of effect between the two variables: if you are making a prediction that there is a negative or positive relationship between the variables, then the one-tailed test is appropriate; if you are not making a directional prediction, you should use the two-tailed test if there is not a specific prediction about the direction of the relationship between the variables you are correlating. The selections in the above dialog box will produce the following output:

Correlations

		Educational Level (years)	Current Salary	Previous Experience (months)
Educational Level (years)	Pearson Correlation	1.000	.661**	-.262*
	Sig. (2-tailed)	.	.000	.000
	N	474	474	474
Current Salary	Pearson Correlation	.661**	1.000	-.007*
	Sig. (2-tailed)	.000	.	.034
	N	474	474	474
Previous Experience (months)	Pearson Correlation	-.262**	-.007*	1.000
	Sig. (2-tailed)	.000	.034	.
	N	474	474	474

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

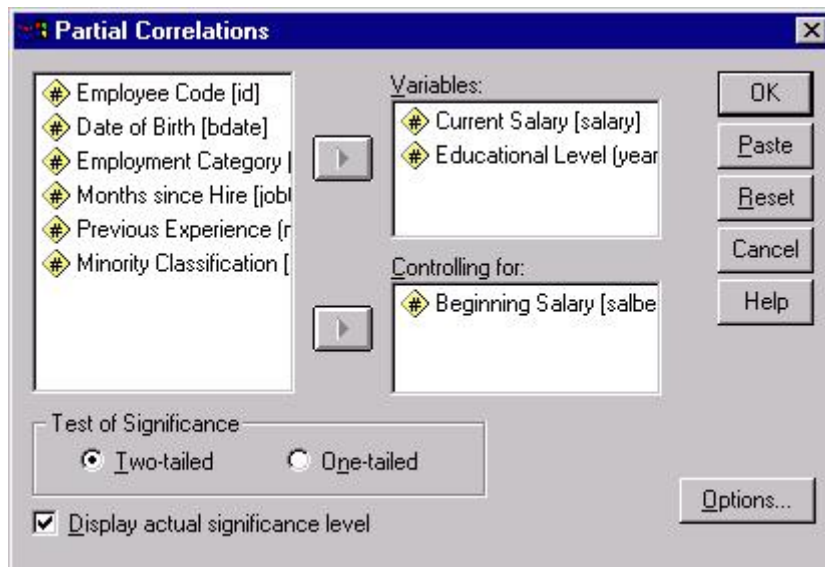
This output gives us a correlation matrix for the three correlations requested in the above dialog box. Note that despite there being nine cells in the above matrix, there are only three correlation coefficients of interest: (1) the correlation between current salary and educational level, the correlation between previous experience and educational level, and the correlation between current salary and previous experience. The reason only three of the nine correlations are of interest is because the diagonal consists of correlations of each variable with itself, always resulting in a value of 1.00 and the values on each side of the diagonal replicate the values on the opposite side of the diagonal. For example, the three unique correlation coefficients show there is a positive correlation between employees' number of years of education and their current salary. This positive correlation coefficient (.661) indicates that there is a statistically significant ($p < .001$) linear relationship between these two variables

such that the more education a person has, the larger that person's salary is. Also observe that there is a statistically significant ($p < .001$) negative correlation coefficient (-.252) for the association between education level and previous experience, indicating that the linear relationship between these two variables is one in which the values of one variable decrease as the other increases. The third correlation coefficient (-.097) also indicates a negative association between employee's current salaries and their previous work experience, although this correlation is fairly weak.

The second type of correlation listed under the *Correlate* menu item is the partial correlation, which measures an association between two variables with the effects of one or more other variables factored out. To obtain a partial correlation, select the following menu item:

Analyze
Correlate
Partial...

This will produce the following dialog box:



Here, we have selected the variables we want to correlate as well as the variable for which we want to control by first clicking on variable names to highlight them on the left side of the box, then moving them to the boxes on the right by clicking on the arrow immediately to the left of either the *Variables* box or the *Controlling for* box. In this example, we are correlating current salaries with years of education while controlling for beginning salaries. Thus, we will have a measure of the association between current salaries and years of education, while removing the association between beginning salaries and the two variables we are correlating. The above example will produce the following output:

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S - - -

Controlling for.. SALBEGIN

	SALARY	EDUC
SALARY	1.0000 (0) P= .	.2810 (471) P= .000
EDUC	.2810 (471) P= .000	1.0000 (0) P= .

(Coefficient / (D.F.) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed

Notice that the correlation coefficient is considerably smaller in the output above than in the bivariate correlation example: the correlation between these variables was .661, whereas it is only .281 in the partial correlation. Nevertheless, a statistically significant association ($p < .001$) exists between these variables. While it may be obvious in the above example that starting and current salaries will be highly correlated, the example illustrates how partial correlations can be used to assess the extent to which variables can be used to explain unique variance by removing the effects of other variables that may be highly correlated with the relationship of interest.

Partial correlations can be especially useful in situations where it is not obvious whether variables possess a unique relationship or whether several variables overlap with each other. For example, if you were attempting to correlate anxiety with job performance and stress with job performance, it would be useful to conduct partial correlations. You could correlate anxiety and a job performance measure while controlling for stress to determine if there were a unique relationship between anxiety and job performance or whether perhaps stress is highly correlated with anxiety--which would result in little remaining variance that could be uniquely attributed to the association between anxiety and job performance.

Regression

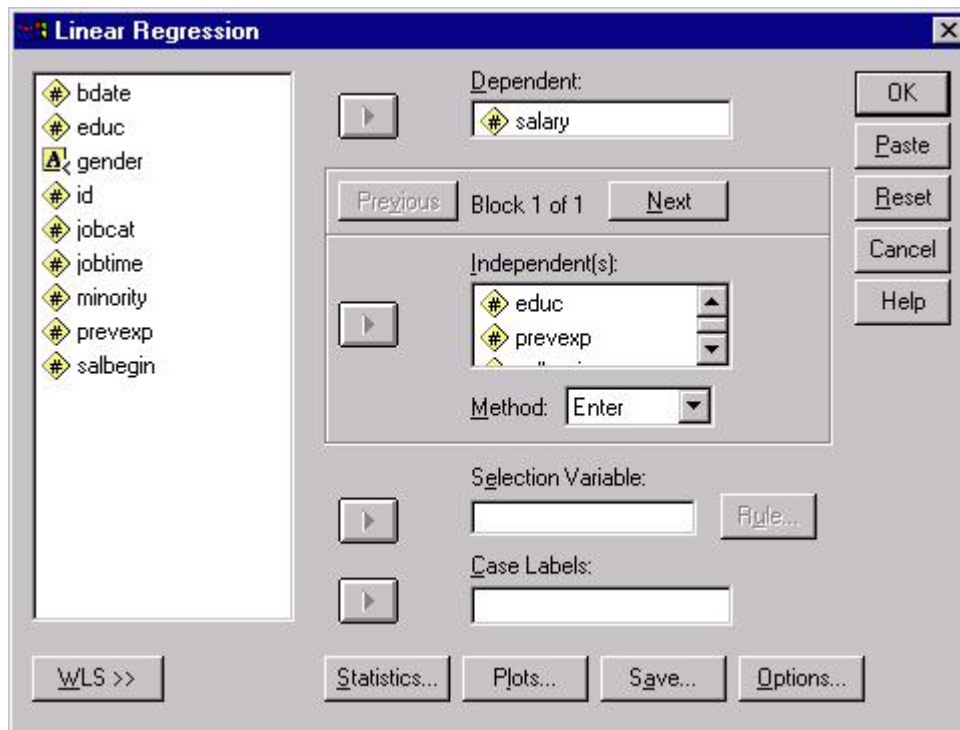
Regression is a technique that can be used to investigate the effect of one or more predictor variables on an outcome variable. Regression allows you to make statements about how well one or more independent variables will predict the value of a dependent variable. For example, if you were interested in investigating which variables in the employee database were good predictors of employees' current salaries, you could create a regression equation that would use several of the variables in the dataset to predict employees' salaries. By doing this you will be able to make statements about whether knowing something about variables such as

employees' number of years of education, their starting salary, or their number of months on the job are good predictors of their current salaries.

To conduct a regression analysis, select the following from the *Analyze* menu:

Analyze
Regression
Linear...

This will produce the following dialog box:



This dialog box illustrates an example regression equation. As with other analyses, you select variables from the box on the left by clicking on them, then moving them to the boxes on the right by clicking the arrow next to the box where you want to enter a particular variable. Here, employees' current salary has been entered as the dependent variable. In the *Independent(s)* box, several predictor variables have been entered, including education level, beginning salary, months since hire, and previous experience.

NOTE: Before you run a regression model, you should consider the method that you use for selecting or rejecting variables in that model. The box labeled *Method* allows you to select from one of five methods: *Enter*, *Remove*, *Forward*, *Backward*, and *Stepwise*. Unfortunately, we cannot offer a comprehensive discussion of the characteristics of each of these methods here, but you have several options regarding the method you use to remove and retain predictor variables in your regression equation. In this example, we will use the SPSS default method, *Enter*, which is a standard approach in regression models. If you have questions

about which method is most appropriate for your data analysis, consult a regression text book, the SPSS help facilities, or contact a consultant.

The following output assumes that only the default options have been requested. If you have selected options from the *Statistics*, *Plots*, or *Options* boxes, then you will have more output than is shown below and some of your tables may contain additional statistics not shown here.

The first table in the output, shown below, includes information about the quantity of variance that is explained by your predictor variables. The first statistic, *R*, is the multiple correlation coefficient between all of the predictor variables and the dependent variable. In this model, the value is .90, which indicates that there is a great deal of variance shared by the independent variables and the dependent variables. The next value, *R Square*, is simply the squared value of *R*. This is frequently used to describe the goodness-of-fit or the amount of variance explained by a given set of predictor variables. In this example, the value is .81, which indicates that 81% of the variance in the dependent variable is explained by the independent variables in the model.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.900 ^a	.810	.809	\$7,465.14

a. Predictors: (Constant), Months since Hire, Previous Experience (months), Beginning Salary, Educational Level (years)

The second table in the output is an ANOVA table that describes the overall variance accounted for in the model. The *F* statistic represents a test of the null hypothesis that the expected values of the regression coefficients are equal to each other and that they equal zero. Put another way, this *F* statistic tests whether the *R square* proportion of variance in the dependent variable accounted for by the predictors is zero. If the null hypothesis were true, then that would indicate that there is not a regression relationship between the dependent variable and the predictor variables. But, instead, it appears that the four predictor variables in the present example are not all equal to each other and could be used to predict the dependent variable, current salary, as is indicated by a large *F* value and a small significance level.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	111779919524.3	4	27944379881.0 ^a	501.430	.000 ^a
	Residual	281385759 ^a 2.07	489	55728306.849		
	Total	1137916495436.3	473			

a. Predictors: (Constant), Educational Level (years), Months since Hire, Previous Experience (months), Beginning Salary

b. Dependent Variable: Current Salary

The third table in standard regression output provides information about the effects of individual predictor variables. Generally, there are two types of information in the *Coefficients* table: coefficients and significance tests. The coefficients indicate the increase in the value of the dependent variable for each unit increase in the predictor variable. For example, the unstandardized coefficient for *Educational Level* in the example is 669.91, which indicates to us that for each year of education, a person's predicted salary will increase by \$669.91. A well known problem with the interpretation of unstandardized coefficients is that their values are dependent on the scale of the variable for which they were calculated, which makes it difficult to assess the relative influence of independent variables through a comparison of unstandardized coefficients. For example, comparing the unstandardized coefficient of *Education Level*, 669.91, with the unstandardized coefficient of the variable *Beginning Salary*, 1.77, it could appear that Educational Level is a greater predictor of a person's current salary than is *Beginning Salary*. We can see that this is deceiving, however, if we examine the standardized coefficients, or *Beta coefficients*. Beta coefficients are based on data expressed in standardized, or z score form. Thus, all variables have a mean of zero and a standard deviation of one and are thus expressed in the same units of measurement. Examining the Beta coefficients for *Education Level* and *Beginning Salary*, we can see that when these two variables are expressed in the same scale, *Beginning Salary* is more obviously the better predictor of *Current Salary*.

In addition to the coefficients, the table also provides a significance test for each of the independent variables in the model. The significance test evaluates the null hypothesis that the unstandardized regression coefficient for the predictor is zero when all other predictors' coefficients are fixed to zero. This test is presented as a *t* statistic. For example, examining the *t* statistic for the variable, *Months Since Hire*, you can see that it is associated with a significance value of .000, indicating that the null hypothesis, that states that this variable's regression coefficient is zero when all other predictor coefficients are fixed to zero, can be rejected.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-18149.7	3255.478		-4.961	.000
	Months since Hire	181.488	31.248	.095	4.715	.000
	Previous Experience (months)	-7.303	3.528	-.108	-4.904	.000
	Beginning Salary	1.768	.058	.815	30.111	.000
	Educational Level (years)	669.914	185.596	.113	4.045	.000

a. Dependent Variable: Current Salary

General Linear Model

The majority of procedures used for conducting analysis of variance (ANOVA) in SPSS can be found under the *General Linear Model* (GLM) menu item in the *Analyze* menu. Analysis of variance can be used in many situations to determine whether there are differences between

SPSS for Windows: Displaying Data

- [Section 6: Displaying Data](#)
 - [Tables](#)
 - [Exporting Tables in SPSS](#)
 - [Bar Graphs](#)
 - [Scatterplots](#)
 - [Modifying and Exporting Graphs](#)
 - [Interactive Charts](#)

This document is the third module of a four module tutorial series. This document describes the use of SPSS to create and modify tables which can be exported to other applications. Graphical displays of data are also discussed, including bar graphs and scatterplots as well as a discussions on how to modify graphs using the SPSS Chart Editor and Interactive Graphs. If you are not familiar with SPSS or need more information about how to get SPSS to read your data, consult the first module, [SPSS for Windows: Getting Started](#), of this four part tutorial.

Some users prefer to use keystrokes to navigate through SPSS. Information on common keystrokes are available in our [SPSS 10 for Windows Keystroke Manual](#). This set of documents uses a sample dataset, *Employee data.sav*, that SPSS provides. It can be found in the root SPSS directory. If you installed SPSS in the default location, then this file will be located in the following location: C:\Program Files\SPSS\Employee Data.sav.

Section 6: Displaying Data

Tables

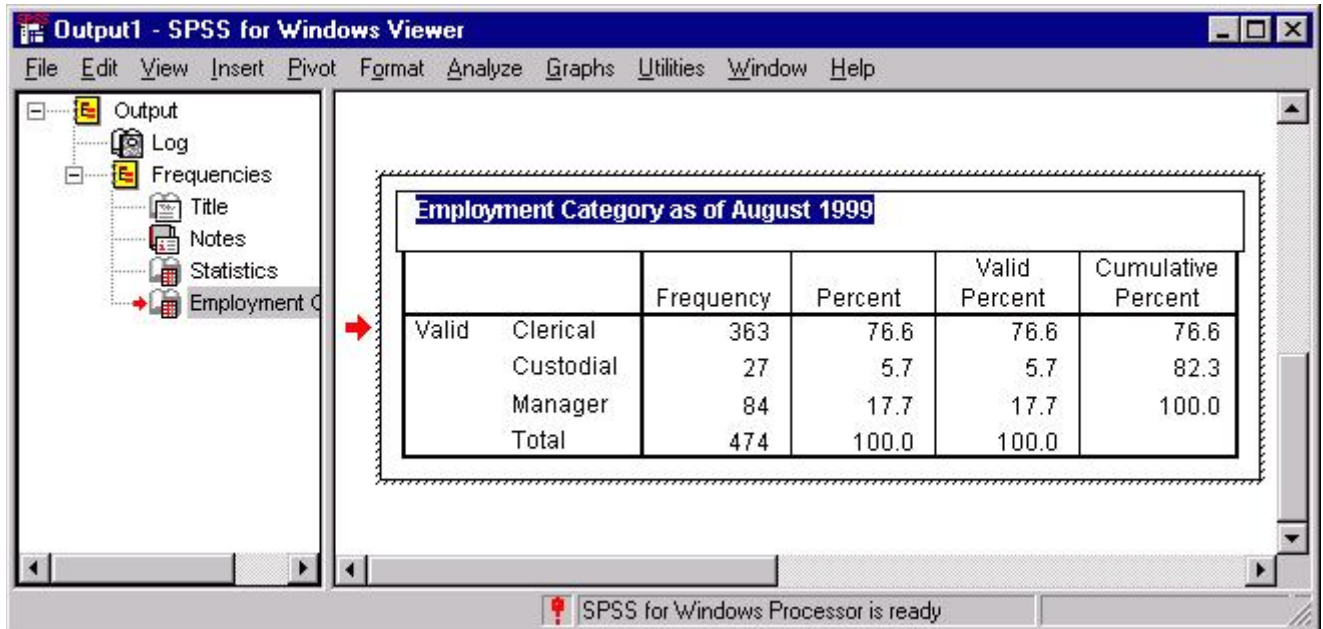
Much of the output in SPSS is displayed in a pivot table format. While these pivot tables are professional quality in their appearance, you may wish to alter their appearance or export them to another application. There are several ways in which you can modify tables. In this section, we will discuss how you can alter text, modify a table's appearance, and export information in tables to other applications.

To edit the text in any SPSS output table, you should first double-click that table. This will outline dashed lines, as shown in the figure below, indicating that it is ready to be edited. Some of the most commonly used editing techniques are changing the width of rows and columns, altering text, and moving text. Each of these topics are discussed below:

- **Changing column width and altering text.** To change column widths, move the mouse arrow above the lines defining the columns until the arrow changes to a double-headed arrow facing left and right. When you see this new arrow, press down on your left mouse button, then drag the line until the column is the width you want, then release your mouse button.
- **Editing text.** First double-click on the cell you wish to edit, then place your cursor on that cell and modify or replace the existing text. For example, in the frequency table

shown below, the table was double-clicked to activate it, then the pivot table's title was double-clicked to activate the title. The original title, "Employment Category," was modified by adding the additional text, "as of August 1999."

- **Using basic editing commands, such as *cut*, *copy*, *delete*, and *paste*.** When you cut and copy rows, columns, or a combination of rows and columns by using the *Edit* menu's options, the cell structure is preserved and these values can easily be pasted into a spreadsheet or table in another application.

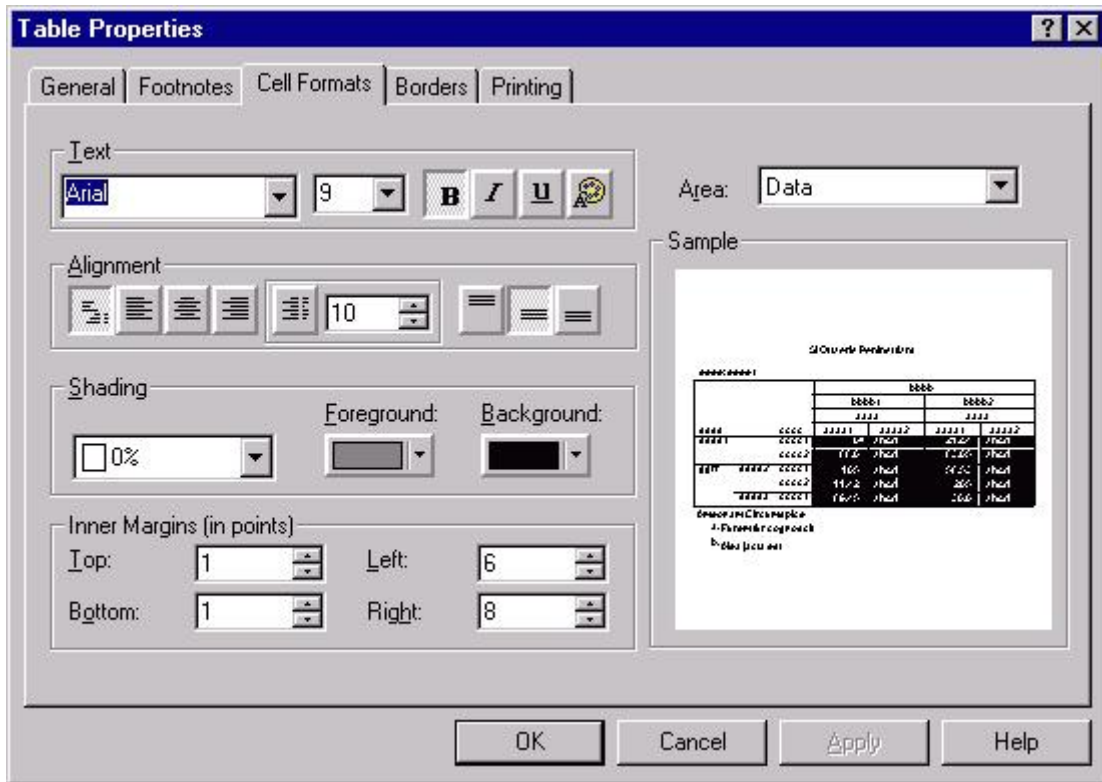


The screenshot shows the 'Output1 - SPSS for Windows Viewer' window. On the left, a tree view shows the hierarchy: Output > Log > Frequencies > Title > Notes > Statistics > Employment C. A red arrow points from the 'Employment C' item to the pivot table displayed in the main window. The pivot table has a title 'Employment Category as of August 1999' and the following data:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clerical	363	76.6	76.6	76.6
	Custodial	27	5.7	5.7	82.3
	Manager	84	17.7	17.7	100.0
	Total	474	100.0	100.0	

[D](#)

Aside from changing the text in a table, you may also wish to change the appearance of the table itself. But first, it is best to have an understanding of the SPSS *TableLook* concept. A TableLook is a file that contains all of the information about the formatting and appearance of a table, including fonts, the width and height of rows and columns, coloring, etc. There are several predefined TableLooks that can be viewed by first right-clicking on an active table, then selecting the *TableLooks* menu item. Doing so will produce the following dialog box:



The above figure shows the *Table Properties* dialog box with the **Cell Formats** tab selected. You can alternate between tabs (e.g., *General*, *Footnotes*, etc.) by clicking on the tab at the upper left of the dialog box. While a complete description of the options available in the *Table Properties* dialog box is beyond the scope of this document, there are a few key concepts that are worth mentioning. Note the *Area* box at the upper right of the dialog box. This refers to the portion of the box that is being modified by the options on the left side of the box. For example, the color in the *Background* of the *Data* portion of the table was changed to black and the color of the text was changed to white by first choosing *Data* from the *Area* box, then selecting black from the *Background* drop-down menu and selecting white for the text by clicking on the color palette icon in the *Text* area on the left side of the dialog box.

The *Printing* tab also has some useful options. For example, the default option for three-dimensional tables containing several layers is that only the visible layer will be printed. One of the options under the *Printing* tab allows you to request that all layers be printed as individual tables. Another useful *Printing* option is the *Rescale wide/long tables to fit page*, which will shrink a table that is larger than a page so that it will fit on a single page.

Any modifications to a specific table can be saved as a TableLook. By saving a TableLook, you will be saving all of the layout properties of that table and can thus apply that look to other tables in the future. To save a TableLook, click on the **General** tab in the *Table Properties* dialog box. There are three buttons on the bottom right of this box. Use the **Save Look** button to save a TableLook. That button will produce a standard *Save As* dialog box with which you can save the TableLook you created.

Exporting Tables in SPSS

In addition to modifying a table's appearance, you may also wish to export that table. There are three primary ways to export tables in SPSS. To get a menu that contains the available options for exporting tables, right-click on the table you wish to export. The three options for exporting tables are: *Copy*, *Copy object*, and *Export*.

The *Copy* option copies the text and preserves the rows and columns of your table but does not copy formatting, such as colors and borders. This is a good option if you want to modify the table in another application. When you select this option, the table will be copied into your system clipboard. Then, to paste the table, select the *Paste* command from the *Edit* menu in the application to which you are importing the table. The *Copy* option is useful if you plan to format your table in the new application; the disadvantage of this method is that only the text and table formatting remains and you will therefore lose much of the formatting that you observe in the Output Viewer.

The *Copy object* method will copy the table exactly as it appears in the SPSS Output Viewer. When you select this option, the table will be copied into your clipboard and can be imported into another application by selecting the *Paste* option from the *Edit* menu of that application. When you paste the table using this option, it will appear exactly as it is in the Output Viewer. The disadvantage of this method is that it can be more difficult to change the appearance of the table once it has been imported.

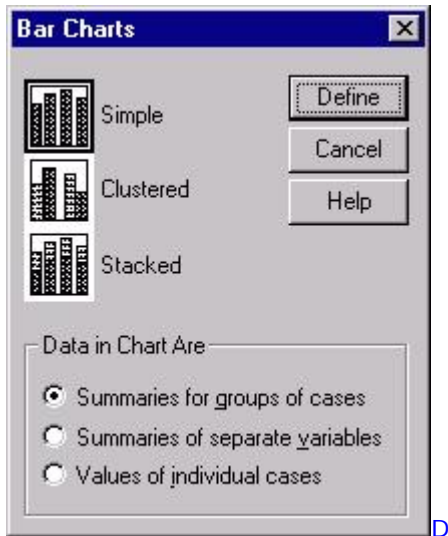
The third method, *Export*, allows you to save the table as an HTML or an ASCII file. The result is similar to the the *Copy* command: you will have a table that retains the text and cell layout of the table you exported, but it will retain little formatting. This method for exporting tables to other applications is different from the above two methods in that it creates a file containing the table rather than placing a copy in the system clipboard. When you select this method, you will immediately be presented with a dialog box allowing you to choose the format of the file you are saving and its location on disk. The primary advantage of this method is that you can immediately create an HTML file that can be viewed in a Web browser.

Bar Graphs

Bar graphs are a common way to graphically display the data that represent the frequency of each level of a variable. They can be used to illustrate the relative frequency of levels of a variable by graphically displaying the number of cases at each level of the variable. For example, you could compare the distribution of employees among employment categories by plotting the frequency of each job category next to the other job categories. If you want to develop a similar display for a continuous variable, examine the *Histogram* option on the *Graphs* menu. To access the SPSS facilities for visually displaying data, including the bar graph procedure, select the *Bar* option from the *Graphs* menu:

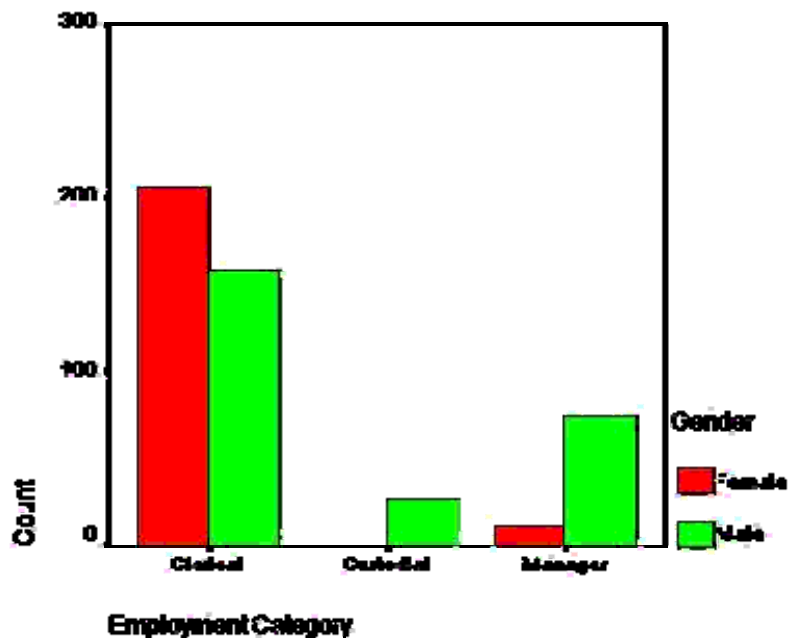
Graphs
Bar...

This will produce the following dialog box:

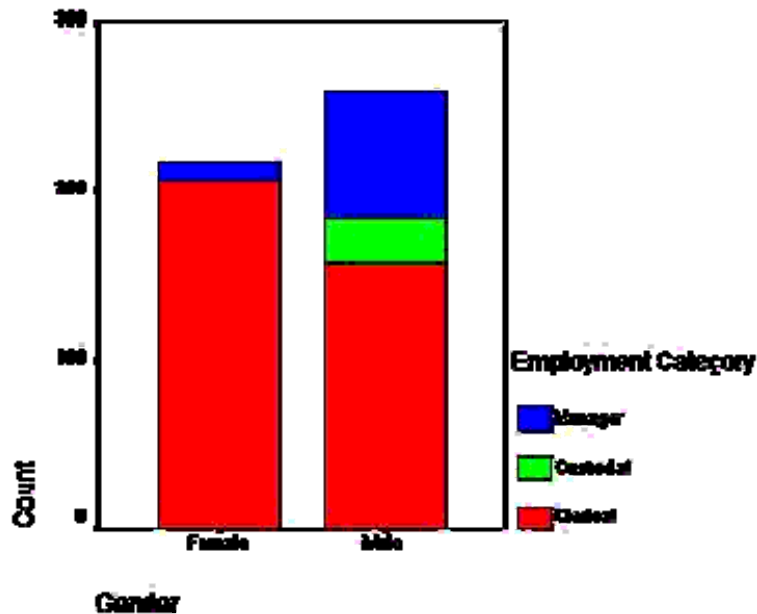


In this dialog box, you must choose between the three types of bar graphs offered by SPSS. The *Simple* bar graph is the most common one and is used to graph frequencies of levels within a variable. In the *Simple* bar graph, the category axis is the variable and each bar represents a level of the variable.

The other two types of bar graphs are used in situations where you want to graph frequencies for more than one variable. For example, if you want to graph job category by gender of employee, then you would choose either the *Clustered* or *Stacked* options in the above dialog box. The *Clustered* option groups together bars representing levels of a category. For example, in graphing gender by job category, you could produce a bar graph in which male and female, the levels of the gender variable, are grouped next to each other for each of the three levels of the job-category variable:

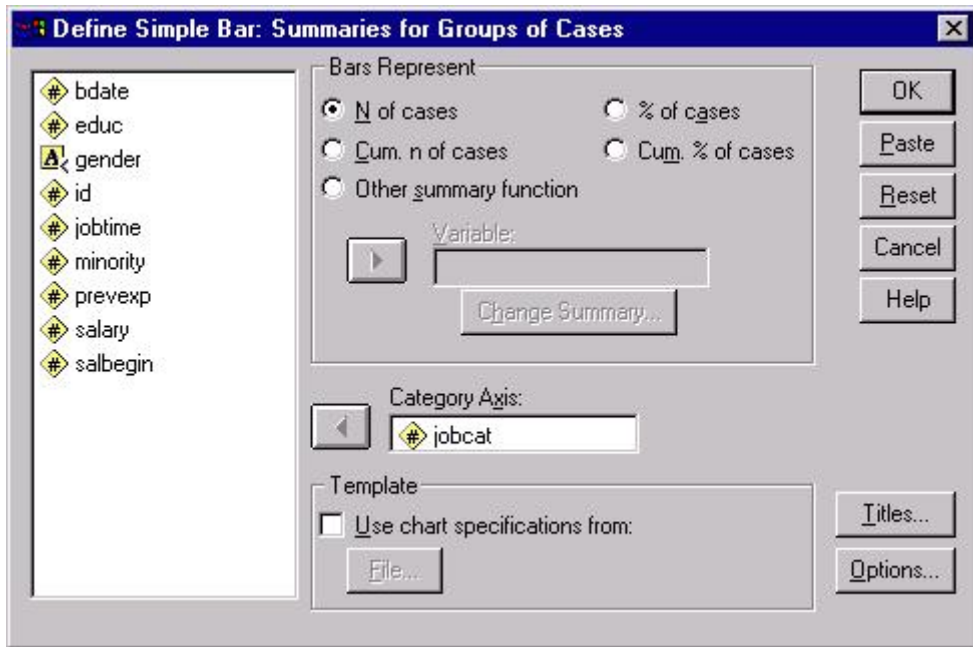


The other bar graph option for two variables is the *Stacked* option, which will have a bar for each level of one of the variables, while the levels of the other variable are placed on top of each other within each bar. The frequency of each level within a variable in a stacked bar graph is represented by a different color. For example, you could create a graph in which there was a bar for the levels of gender, male and female, and the numbers of males in females within each employment category are represented by a different color within a bar:



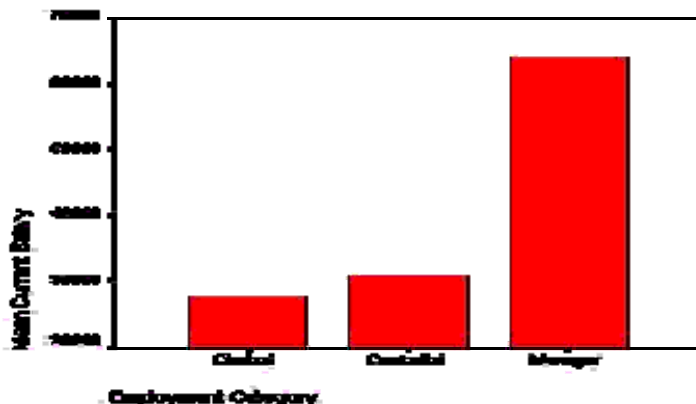
[D](#)

This discussion will focus on the the *Simple* graph. The options for all three are similar, so an understanding of the *Simple* option will suffice for either the *Clustered* or *Stacked* options. To get started with the bar graph, click on the icon representing the type of graph that you want, then click on the **Define** button to produce the following dialog box:



In this dialog box, the *Category Axis* is the only box that you must fill in. Click on a variable name in the list of variables at left and move it to the *Category Axis* box by clicking on the arrow next to that box. You should also select the summary information that you want in the *Bars Represent* section of the box. The *N of cases* option is the default and is the most common way of summarizing data in a bar graph. There is little difference between using the *N of cases* and the *% of cases* options, other than the unit of measurement in the vertical axis. Using either the *Cum. n of cases* or the *Cum. % of cases* options will produce a cumulative graph in which the first bar represents the level of a variable with the fewest number of cases, then the next bar represents the level with the second fewest number of cases in addition to the level that has previously been graphed, and so on.

In addition to these four options, you can also use the *Other summary function* option to specify another summary statistic, such as a variable's mean, sum, or variance. As in the above example, begin by selecting a *Category Axis* variable. Next, select the *Other summary function*, then place another variable in the box labels *Variable*. For example, you could select the variable *salary* to obtain the mean salary for each level of the variable *jobcat*. Doing so would produce the following table:

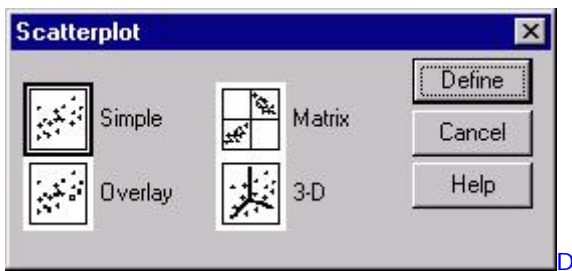


Scatterplots

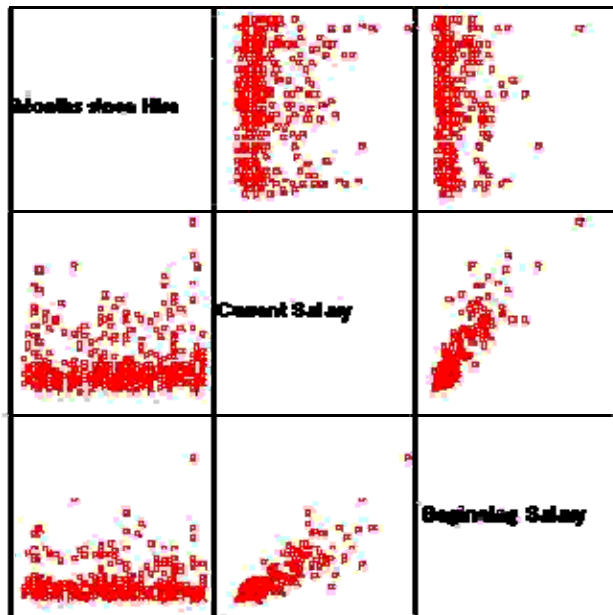
Scatterplots give you a tool for visualizing the relationship between two or more variables. Scatterplots are especially useful when you are examining the relationship between continuous variables using statistical techniques such as correlation or regression. Scatterplots are also often used to evaluate the bivariate relationships in regression analyses. To obtain a scatterplot in SPSS, go to the *Graphs* menu and select *Scatter*:

Graphs Scatter...

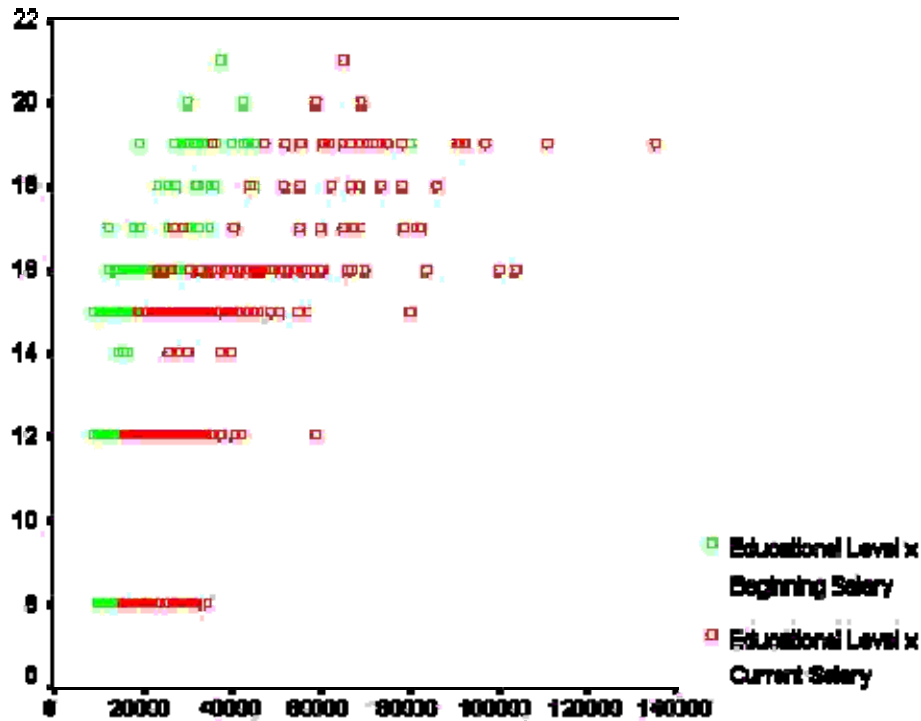
This will produce the following dialog box:



Note that there are four scatterplot options. The *Simple* option graphs the relationship between two variables. The *Matrix* option is for two or more variables that you want graphed in every combination: variable is plotted with every other variable. Every combination is plotted twice so that each variable appears on both the X and Y axis. For example, if you specified a *Matrix* scatterplot with three variables, *salary*, *salbegin*, and *jobtime*, you would receive the following scatterplot matrix:

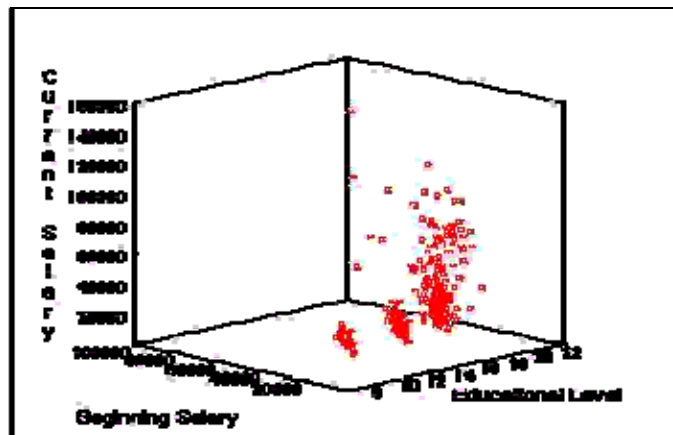


The third scatterplot option is the *Overlay* option. It allows you to plot two scatterplots on top of each other. The plots are distinguished by color on the overlaid plot. For example, you could plot education by beginning and current salaries by pairing the variables *educ* with *salbegin* and *educ* with *salary* using our example dataset. Doing so would produce the graph shown below. In that graph, the green points represent the *educ* x *salbegin* plot whereas the red points represent the *educ* x *salary* plot.



[D](#)

The fourth option for scatterplots is the 3-D scatterplot. This is used to plot three variables in three dimensional space. Here is an example of the 3-D option, containing the variables, *educ*,



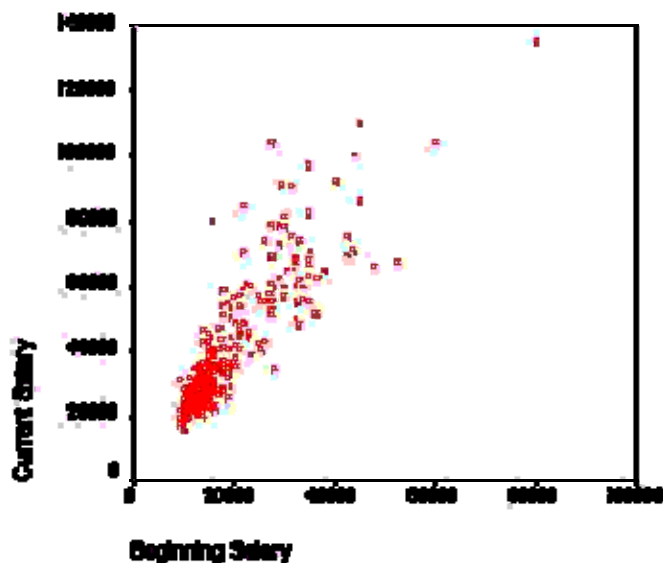
salary, and *salbegin*:

[D](#)

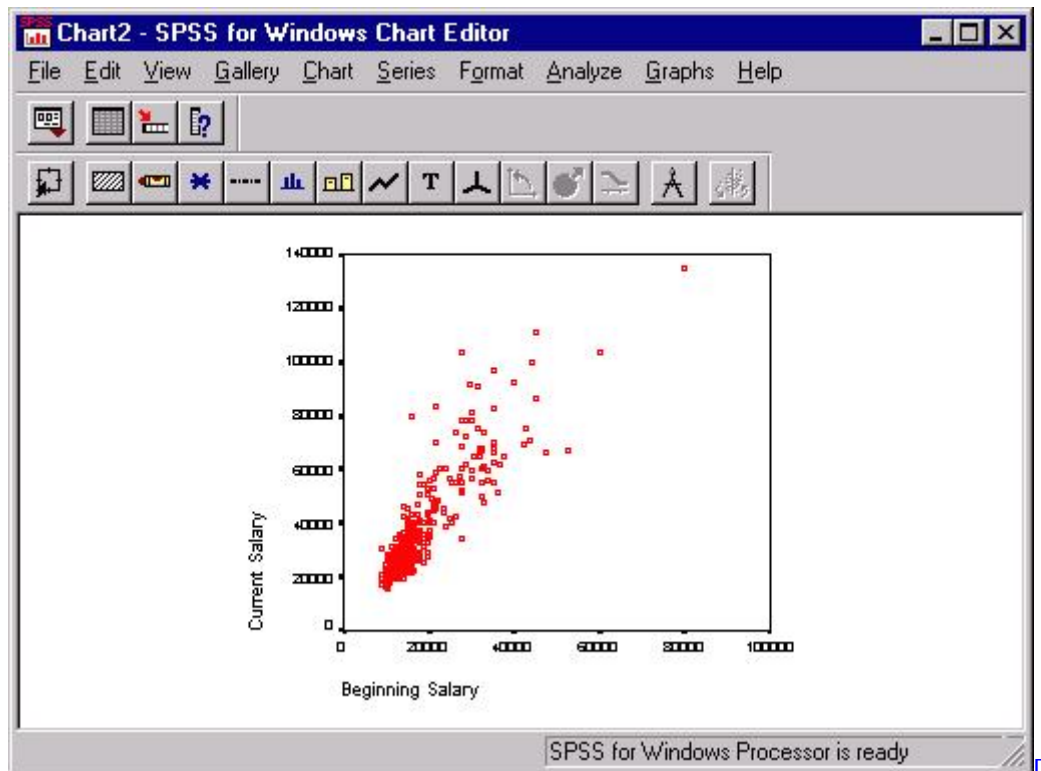
Returning to the *Simple* scatterplot option, you can examine some of the the options that are commonly used with a basic scatterplot by plotting the values of two variables. When you select the *Simple* option from the initial dialog box, you will get the following dialog box:



Select the two variables that you want to plot from the list on the left, placing one in the *Y axis* box and the other in the *X axis* box. If you have multiple groups (e.g., males and females) in your dataset, you can have SPSS draw different colored markers for each group by entering a group variable in the *Set Markers by* box. You can add titles to your chart by clicking on the **Titles** button. Clicking **OK** will produce the following scatterplot:



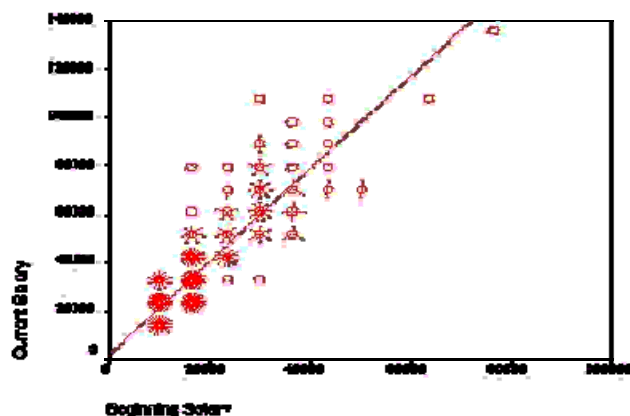
Some of the most useful options for modifying your scatterplot are only available after you have the initial scatterplot created. To use these options, double-click on the chart. This will open the chart in a new window, as shown below:



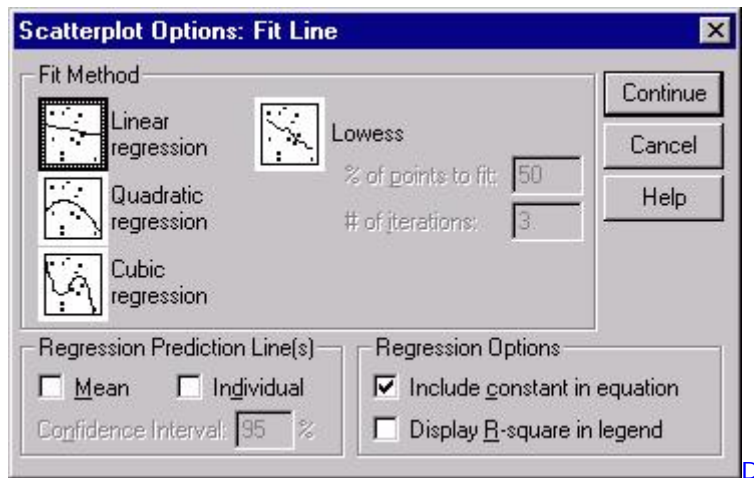
From this window, you have several options for modifying your chart. We will only deal with scatterplot-specific options here, as many of the general options will be covered in the next section. To get the scatterplot options, select *Options* from the *Chart* menu:

Chart Options...

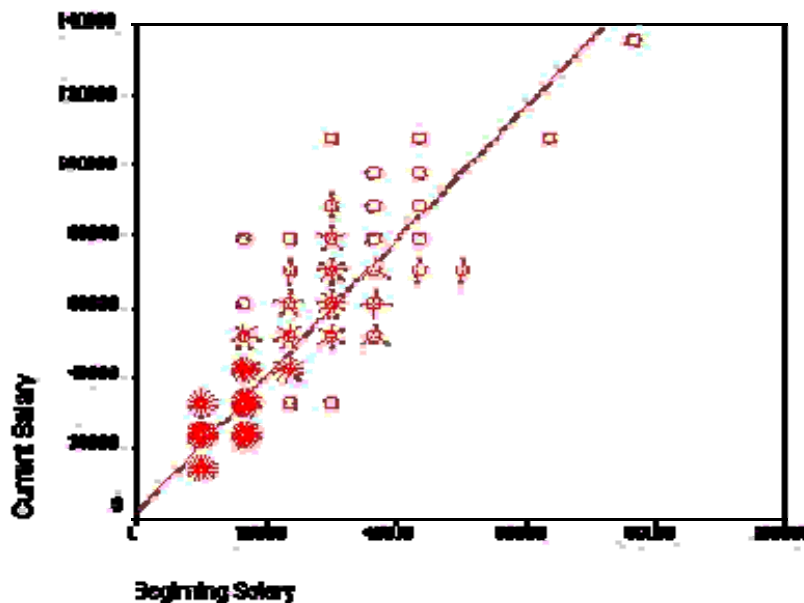
To get the following dialog box:



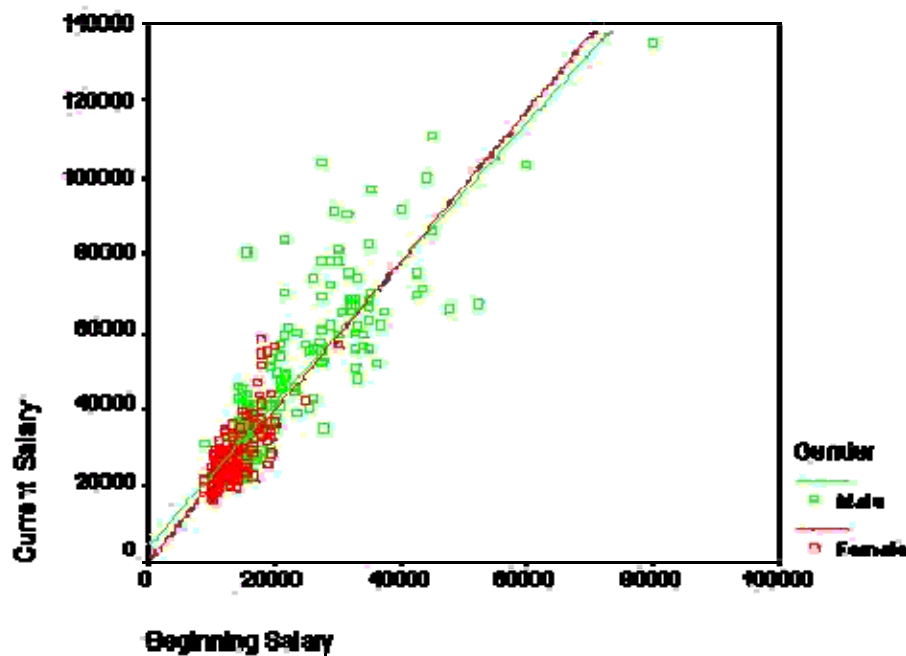
Some of the most useful options that will add information to your scatterplot are the *Sunflowers* and the *Fit Line* options. The *Show sunflowers* option draws spikes that resemble petals for each case where there are overlapping cases in the scatterplot. Clicking on the **Sunflower Options** button will then give you several options for defining how fine or coarse you want the spikes to appear and how many cases are represented by each spike in your scatterplot. See the chart at the end of this section for an example of sunflowers on a scatterplot. The *Fit Line* option will allow you to plot a regression line over your scatter plot. Click on the **Fit Options** button to get this dialog box:



Next, click on the type of regression line you want in the *Fit Method* box. The *Linear regression* line is the most commonly plotted regression line, and it represents the linear increase in the Y variable as a function of increases in the X variable. The graph below illustrates a linear regression line plotted on a sunflower scatterplot. If you have a quadratic or cubic term in your model, you will likely be interested in plotting one of those fit lines.



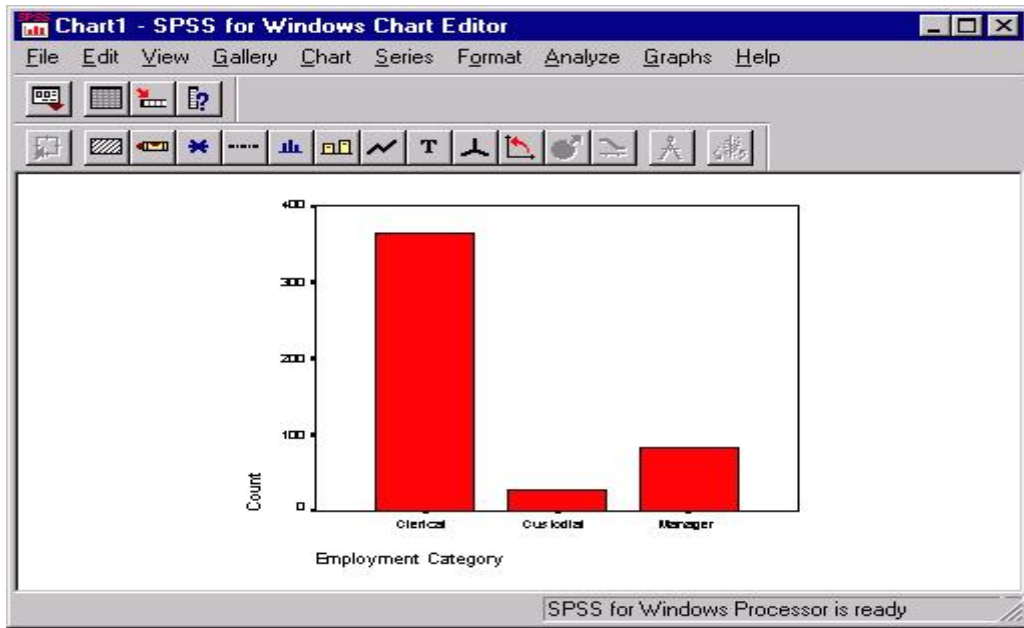
One common extension of the above example is to plot separate regression lines for subgroups. For example, you could plot separate regression lines for men and women in the above example to visually examine whether the relationship between beginning salaries and current salaries is the same for both men and women. To do this, you will first need to select a categorical variable in the main scatterplot dialog box (this is the dialog box that is labeled *Simple Scatterplot*). Place a categorical variable in the box labeled *Set Markers by*. After doing that, follow the rest of the above steps until you reach the dialog box labeled *Scatterplot Options*. There, you can define separate regression lines for your groups by selecting the *Subgroups* option in the *Fit Line* section of the *Scatterplots* dialog box. Doing so produces the graph shown below:



[D](#)

Modifying and Exporting Graphs

The primary tool for modifying charts in SPSS is the *Chart Editor*. The Chart Editor will open in a new window, displaying a chart from your Output Viewer. The Chart Editor has several tools for changing the appearance of your charts or even the type of chart that you are using. To open the Chart Editor, double-click on an existing chart and the Chart Editor window will open automatically. The Chart Editor shown below contains a bar graph of employment categories:

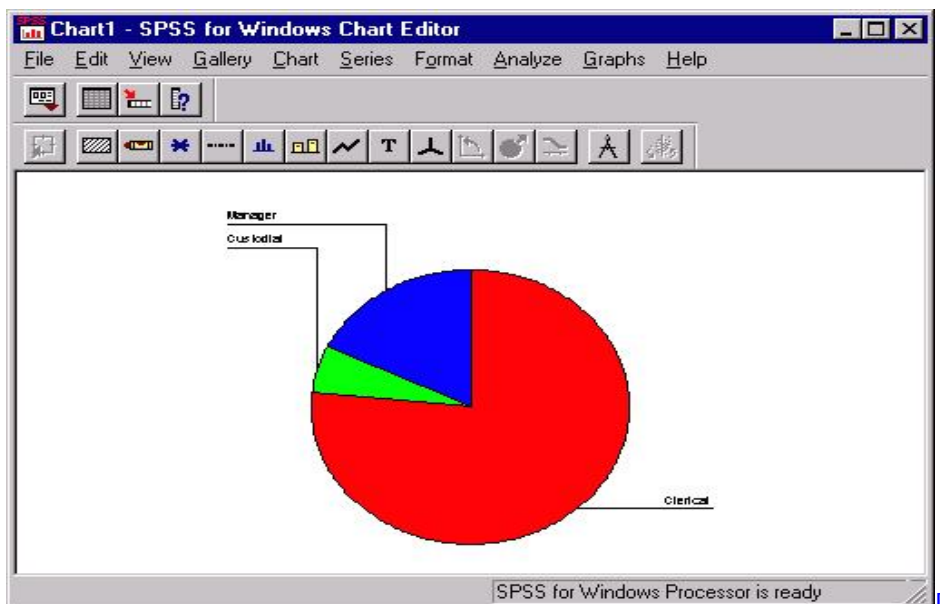


While there are many useful features in the Chart Editor, we will concentrate on the three of them: changing the type of chart, modifying text in the chart, and modifying the graphs.

You can change the type of chart that you are using to display your data using the Chart Editor. For example, if you want to compare how your data would look when displayed as a bar graph and as a pie chart, you can do this from the *Gallery* menu:

Gallery Pie...

Selecting this option will change the above bar graph into the following pie chart:



Once you have selected your graphical look, you can start modifying the appearance of your graph. One aspect of the chart that you may want to alter is the text, including the titles, footnotes, and value labels. Many of these options are available from the *Chart* menu. For example, the *Title* option could be selected from the *Chart* menu to alter the charts title:

Chart Title...

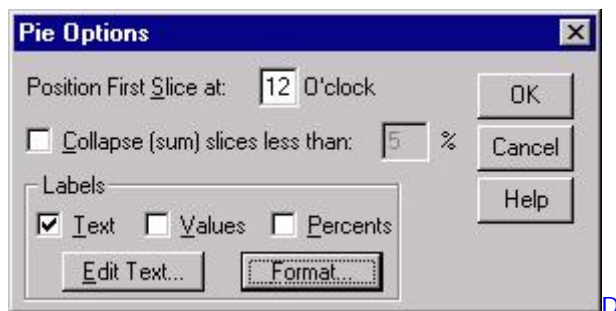
Selecting this menu item will produce the following dialog box:



The title "Employment Categories" was entered in the box above and the default justification was changed from left to center in the *Title Justification* box. Clicking **OK** here would cause this title to appear at the top center of the above pie chart. Other text in the chart, such as footnotes, legends, and annotations, can be altered similarly. The labels for the individual slices of the pies can also be modified, although it may not be obvious from the menu items. To alter the labels for areas of the pie, choose the *Options* item from the *Chart* menu.

Chart Options...

This will produce the following dialog box:



In addition to providing some general options for displaying the slices, the *Labels* section enables you to alter the text labeling slices of the pie chart as well as format that text. You can click the **Edit Text** button to change the text for the labels. Doing so will produce the following dialog box:

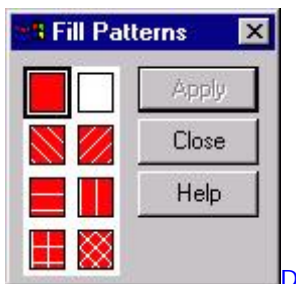


To edit individual labels, first click on the current label, which will be displayed below the *Label* box, then alter the text in the *Label* box. When you finish, click the **Continue** button to return to the *Pie Options* dialog box. You can make changes to the format of your labels by clicking the **Format** button here. If you do not want to change formatting, click on **OK** to return to the Chart Editor.

In addition to altering the text in your chart, you may also want to change the appearance of the graph with which you are working. Options for changing the appearance of graphs can be accessed from the *Format* menu. Many options available from this menu are specific to a particular type of graph. There are some general options that are worth discussing here. One such is *Fill Pattern* option, which changes the pattern of the graph. It can be obtained by selecting the *Fill Pattern* option from the *Format* menu:

Format **Fill Pattern...**

This will produce the following dialog box:



First, click on the portion of the graph where you want to change the pattern, then select the pattern you want by clicking on the pattern sample on the left side of the dialog box. Then, click the **Apply** button to change the appearance of your graph.

One other formatting option that is generally useful is the the ability to change the colors of your graphs. To do that, select the *Color* option from the *Format* menu:

Format **Color...**

To produce the following dialog box:



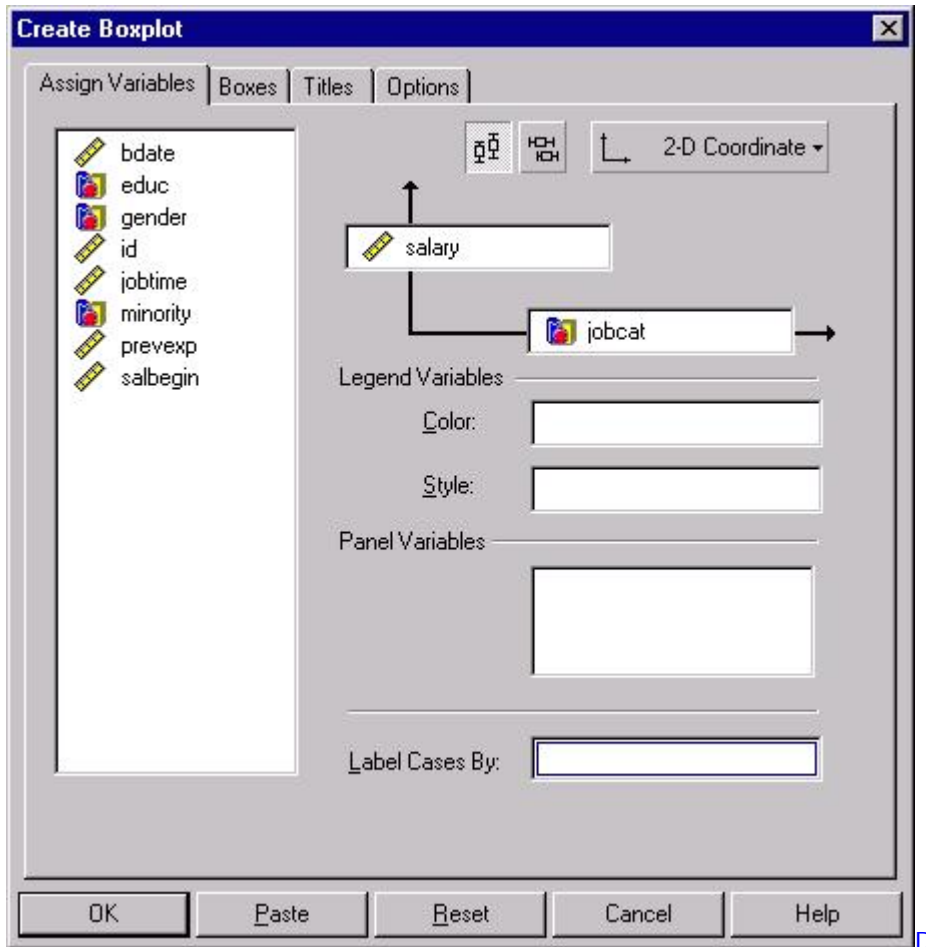
This will allow you to change the color of a portion of a graph and its border. First, select the portion of the graph for which you would like to change its color, then select the *Fill* option if you want to change the color of a portion of the graph and select the *Border* option if you want to change the color of the border for a portion of the graph. Next, click on the color that you want and click **Apply**. Repeat this process for each area or border in the graph that you want to change.

Interactive Charts

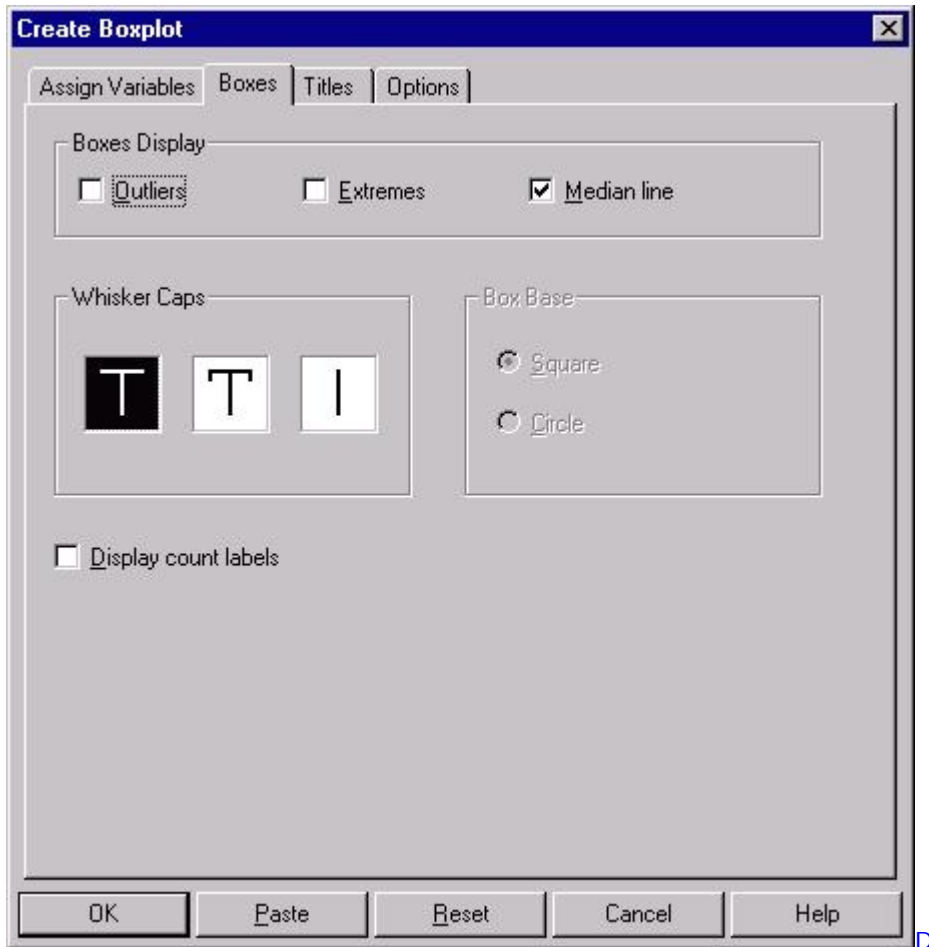
Many of the standard graphs available through SPSS are also available as *interactive charts*. Interactive charts offer more flexibility than standard SPSS graphics: you can add variables to an existing chart, add features to the charts, and change the summary statistics used in the chart. To obtain a list of the available interactive charts, select the *Interactive* option from the *Graphs* menu:

Graphs Interactive

Selecting one of the available options will produce a dialog for designing an interactive graph. For example, if you selected the *Boxplot* option from the menu, you would get this dialog box:

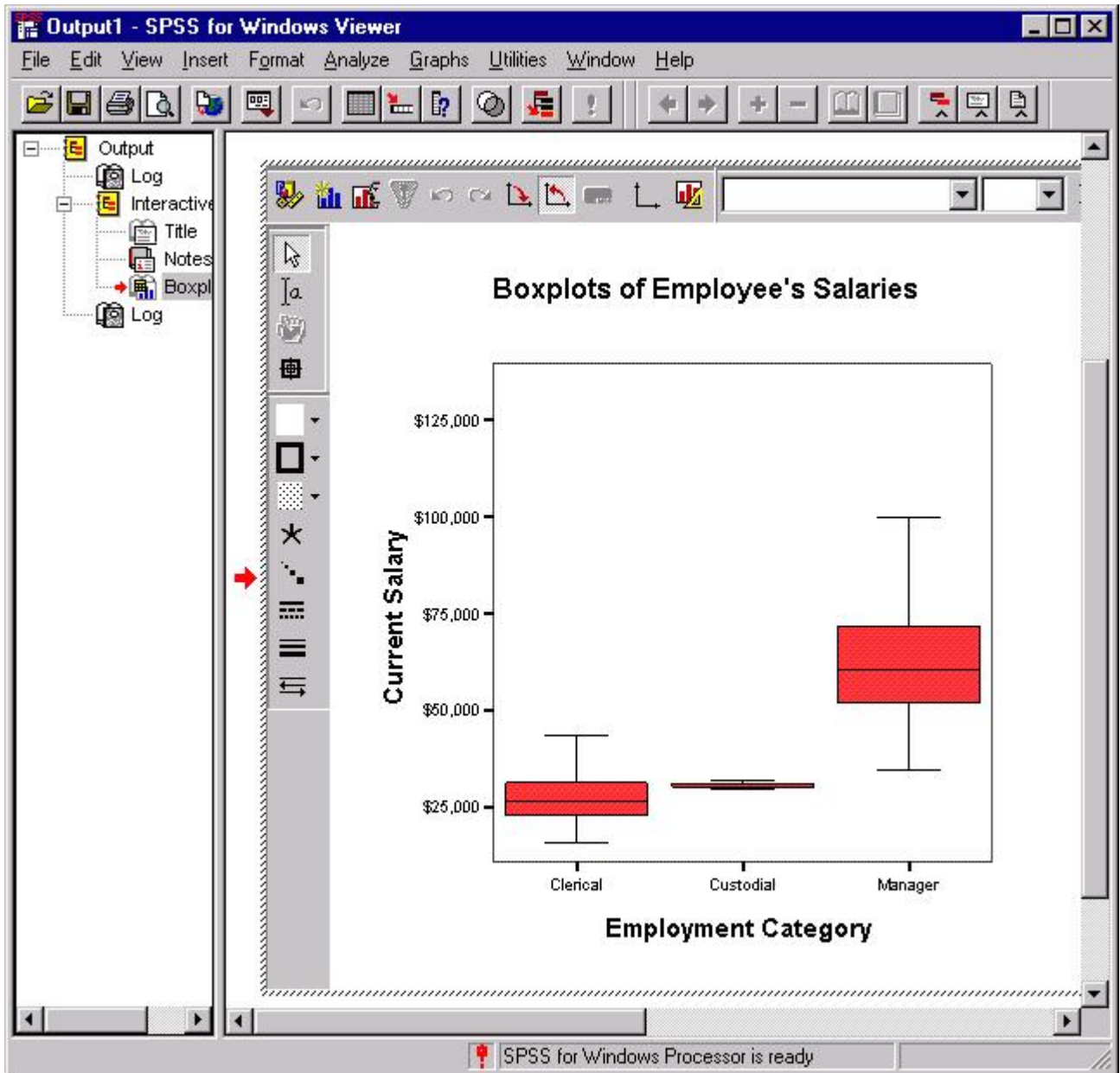


Dialog boxes for interactive charts have many of the same features as other SPSS dialog boxes. For example, in the above dialog box, the variable type is represented by icons: scale variables, such as the variable *bdate*, are represented by the icon that resembles a ruler, while categorical variables, such as the variable *educ*, are represented by the icon that resembles a set of blocks. Variables in the variable list on the left of the dialog box can be moved into the boxes on the right side of the screen by dragging them with your mouse, in contrast to using the arrow button used in other SPSS dialog boxes. Options in interactive graphs can be accessed by clicking on the tabs. For example, clicking on the *Boxes* tab produces the following dialog box:



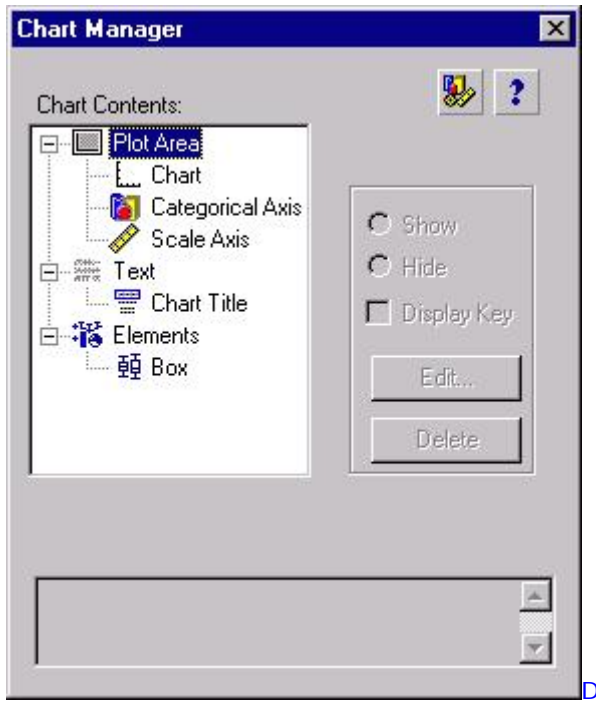
Here, you have several choices about the look of your boxplot. The choice to display the median line is selected here, but the options to indicate outliers and extremes are not selected. The *Titles* and *Options* tabs offer several other choices for altering the look of your table as well, although a thorough discussion of these is beyond the scope of this document. When you have finished the specifications for a graph, click the **OK** button to produce the graph you have specified in the Output Viewer.

Interactive graphs offer several choices for altering the look of the chart after you have a draft in the Output Viewer. To get the menus for doing that, double-click on the interactive graph that you want to alter. For example, double-clicking on the boxplot obtained through the above dialog box will produce the following menus:

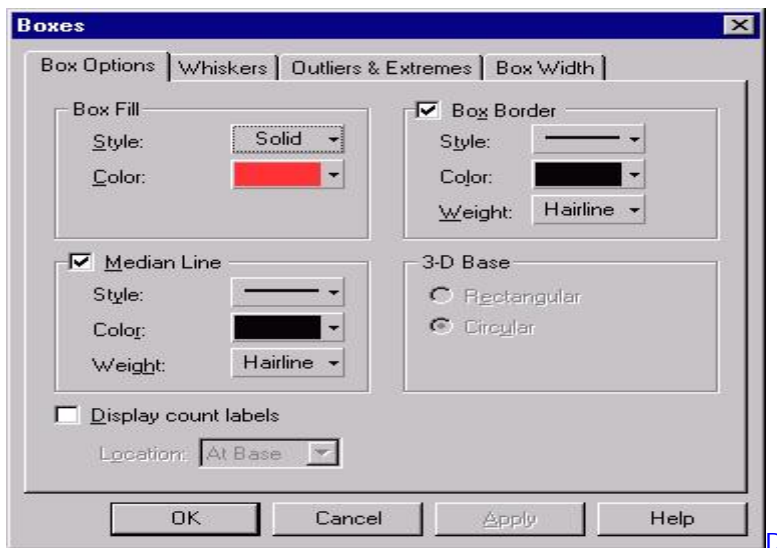


[D](#)

The icons immediately surrounding the graph provide you with several possibilities for altering the look of your graph. The three leftmost items in the horizontal menu are worthy of mention. The leftmost icon produces a dialog box that resembles the original *Interactive Graphs* dialog box and contains many of the same options. For example, you could change the variables that you are graphing using this dialog box. The next icon, the small bar graph, lets you add additional graphical information. For example, you could overlay a line that graphed the means of the three groups in the above graph by choosing the *Dot-Line* option from the menu, or you could add circles representing individuals salaries within each group by choosing the *Cloud* option. The third icon provides several options for changing the look of your chart. Selecting that icon will produce the following dialog box:



Each icon in this dialog box can be double-clicked to produce a dialog box that contains the properties of the component of the chart represented by that icon. For example, you could obtain the properties of the boxes in the interactive graph above by double-clicking on the icon labeled *Box*. Doing so would produce this dialog box:



Changing the properties in this or any other dialog box that controls the properties of any portion of the chart will change the look of the graph in the Output Viewer. For example, you could change the colors of the boxes and their outlines by selecting a different color from the above drop-down menus.