2014

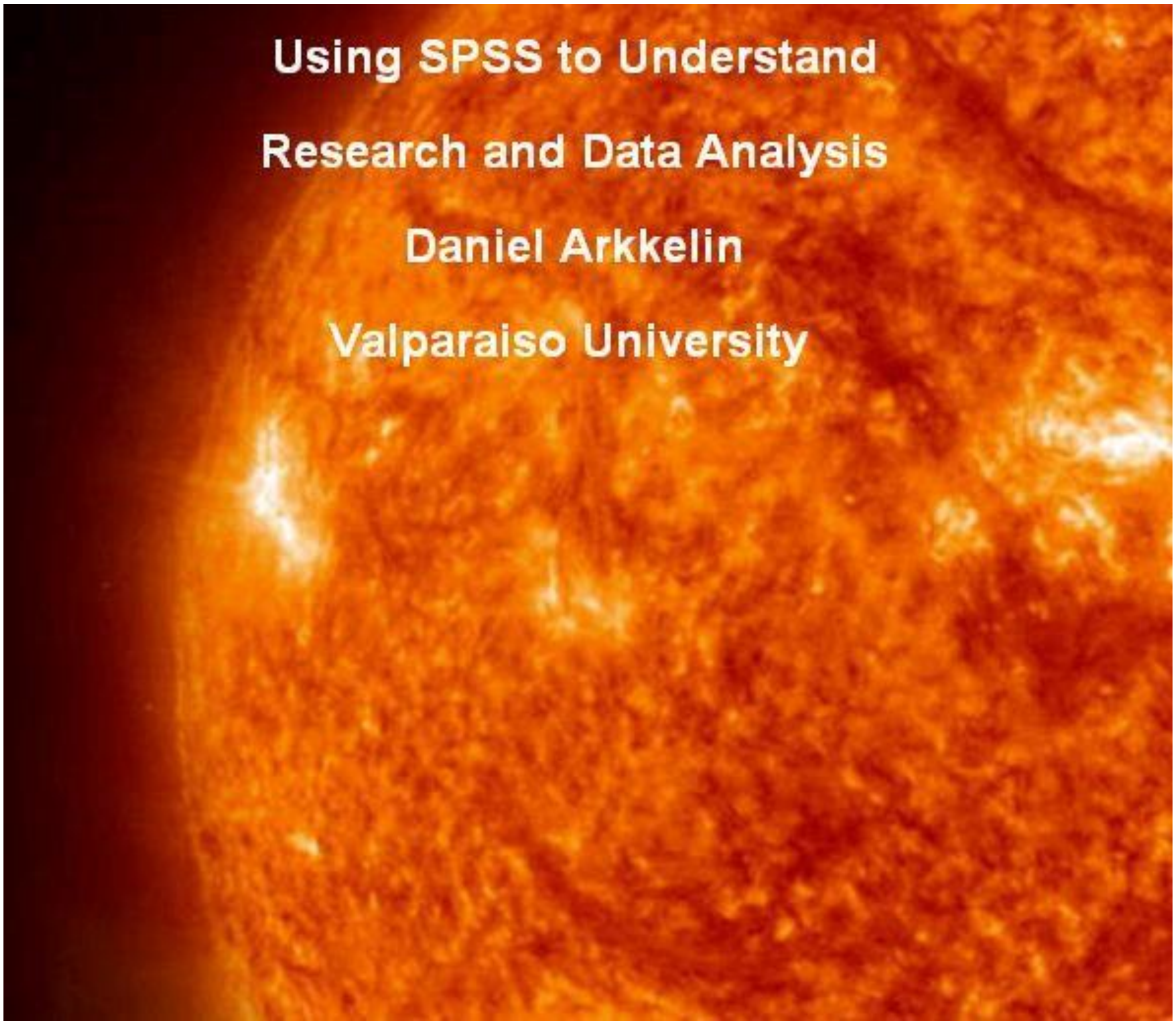
Using SPSS to Understand Research and Data Analysis

Daniel Arkkelin
Valparaiso University, daniel.arkkelin@valpo.edu

2014

Using SPSS to Understand Research and Data Analysis

Daniel Arkkelin
Valparaiso University, daniel.arkkelin@valpo.edu


Using SPSS to Understand
Research and Data Analysis
Daniel Arkkelin
Valparaiso University

Chapter 1
First Contact:
Overview of Book & the SPSS Tutorial

1.1 Goals of this book

We have a number of goals in this book. The first is to provide an introduction to how to use the **Statistical Package for the Social Sciences** (SPSS) for data analysis. The text includes step-by-step instructions, along with screen shots and videos, to conduct various procedures in SPSS to perform statistical data analysis.

However, another goal is to show how SPSS is actually used to understand and interpret the results of research. Thus, we will also discuss the meaning of the data analyses introduced in this text, as well as how a researcher would write these interpretations to present results to others.

This text is designed to be a supplement to undergraduate and graduate courses in statistics and research methods. As such, we do not provide detailed instructions about statistics and methods, leaving that for the respective primary courses.

Although each chapter is designed to be easily accessible to the beginner, we understand that there may be large differences in background with computers, research and statistics. We will assume that you have basic knowledge and experience in these areas. However, if you have minimal experience with computers, research or statistics, we recommend that you read the Appendix, in which we explain some of the basics for the novice.

1.2 Why Have We Chosen to Work with SPSS?

There is no question that business, education, and all fields of science have come to rely heavily on the computer. This dependence has become so great that it is no longer possible to understand social and health science research without substantial knowledge of statistics and without at least some rudimentary understanding of statistical software.

The number and types of statistical software packages that are available continue to grow each year. In this book we have chosen to work with SPSS, or the **Statistical Package for the Social Sciences**. SPSS was chosen because of its popularity within both academic and business circles, making it the most widely used package of its type. SPSS is also a versatile package that allows many different types of analyses, data transformations, and forms of output - in short, it will more than adequately serve our purposes.

The SPSS software package is continually being updated and improved, and so with each major revision comes a new version of that package. In this book, we will describe and use the most recent version of SPSS, called **SPSS for Windows 14.0** Thus, in order to use this text for data analysis, your must have access to the SPSS for Windows 14.0 software package.

However, don't be alarmed if you have an earlier version of SPSS (e.g., Versions 12.0 or 13.0), since the look and feel of SPSS hasn't changed much over the last three versions. Your instructor will explain any differences if you are using a version earlier than 12.0.

1.3 The EZDATA Project and File

Throughout this text we will be illustrating how to compute different statistics in the context of a single, hypothetical research project. Further, we will use the same data file (which we will call **EZDATA**) throughout the book as we demonstrate the various types of data analyses called for by different research methodologies. We believe that this will provide you with a sense of the entire research process, from designing a study, through inputting the data into a file for analysis, to the computation of various statistics and interpretation of the results.

In using the same project and data set throughout, we hope to provide continuity between chapters and give you an appreciation for the unfolding process that researchers experience as they undertake each new analysis of the data. We will introduce this project and the EZDATA file in Chapter 5.

However, to introduce you the main features of SPSS, we will first begin with a simple example using a much smaller data file. That way, you can learn the basics of SPSS procedures before applying them to the more complex EZDATA file, and this will provide added practice before you begin to work on exercises at the end of each chapter (which are likely to be assigned as homework by your instructor).

Once we begin working with the EZDATA file (beginning in Chapter 6), each chapter will include an example procedure using variables in the EZDATA file. The exercises at the end of each chapter will always ask you to complete the same procedures illustrated in the chapter example, but using a different set of variables in the file. This will facilitate your learning and will also provide continuity within and between chapters.

1.4 Suggestions for using this book: Practice Makes Perfect!

This book emphasizes an active learning approach. That is, the best way to learn the skills for using SPSS is to practice them as you are reading about the various procedures introduced in this book. This book is organized as follows:

- We introduce procedures in each chapter by showing actual **Screen Shots** of what you will see in SPSS as each step of the procedure is completed.
- After we explain procedures, we provide short **Show Me! Videos** showing the steps actually being completed in SPSS
- At the end of each chapter we provide **Review Me! Videos** summarizing all of the procedures in that chapter.
- We present **Exercises** at the end of each to give you more practice performing the same procedures in the chapter, but on a different variables or a different example.

Thus, with each example, we explain the steps, show them, review them, and then ask you to apply them to a new example. As you progress through each chapter, we recommend the following:

- **Have SPSS open** on your computer as you read the chapter.
- **Do the examples** yourself by completing each step in SPSS.
- **Save the files** of the chapter examples to help when doing end of chapter exercises.
- **Take notes** on the explanations of the resulting SPSS output files.

Adopting this active approach should solidify your learning how to use SPSS, as well as to help you formulate questions for your instructor if you experience any problems. Another reason to actually do the examples in the chapters is that that your instructor is likely to assign the end-of-chapter exercises as homework!

Thus, if you do the examples in the chapters, the exercises at the end of the chapter will be much easier to do. We believe that following these suggestions will greatly facilitate your learning and understanding of SPSS and interpretation of data analyses in relation to our EZDATA research project.

A word of caution here: we sometimes find that students get confused and turn in the **examples** that they did while reading the chapter rather than the **exercises** at the end of the chapter. So remember:

- The end-of-chapter exercises always involve the same procedures as the in-chapter examples, but on a *different* set of variables of the EZDATA file.
- Although the outputs from the end-of-chapter exercises will look similar to those of the in-chapter examples, the actual variables and statistics will be different.
- Your instructor will most likely be asking you to turn in your output file for the end-of-chapter exercises, not the in-chapter examples.

1.5a The SPSS Start Up Tutorial

SPSS for Windows also has many built-in help features that you can access at any point while using SPSS. We recommend that you run the **SPSS Tutorial**, which you can access when you open up SPSS. You must either have SPSS installed on your hard drive or be able to access SPSS on a network in order to open the program.

If you are confused about how to open SPSS, ask your instructor or see Appendix 1, where we provide more information about accessing SPSS from your hard drive or from a network. For now, to illustrate how to access the SPSS tutorial, we will assume that SPSS is installed on your computer. Regardless of how you access SPSS, the **Start Up Screen** will look the same.

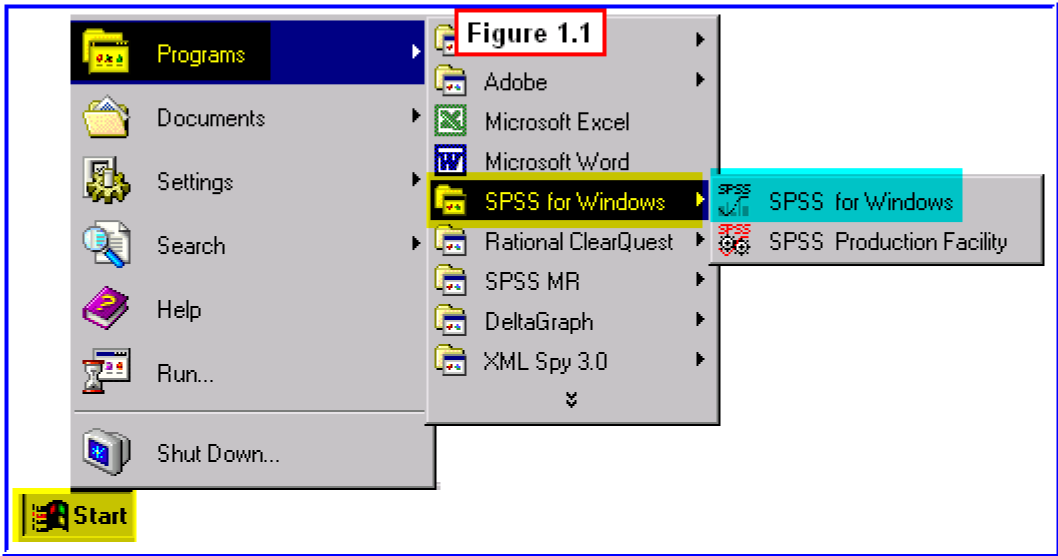
Hint: You have undoubtedly had two programs running simultaneously on your computer. Many people use the **task bar** at the bottom of their computer screen to switch between windows running different programs/subprograms (that is, they alternately click on the boxes on the task bar).

This can be tedious (and sometimes difficult, since many windows can be open simultaneously in SPSS). Here is a handy short-cut for switching between active windows:

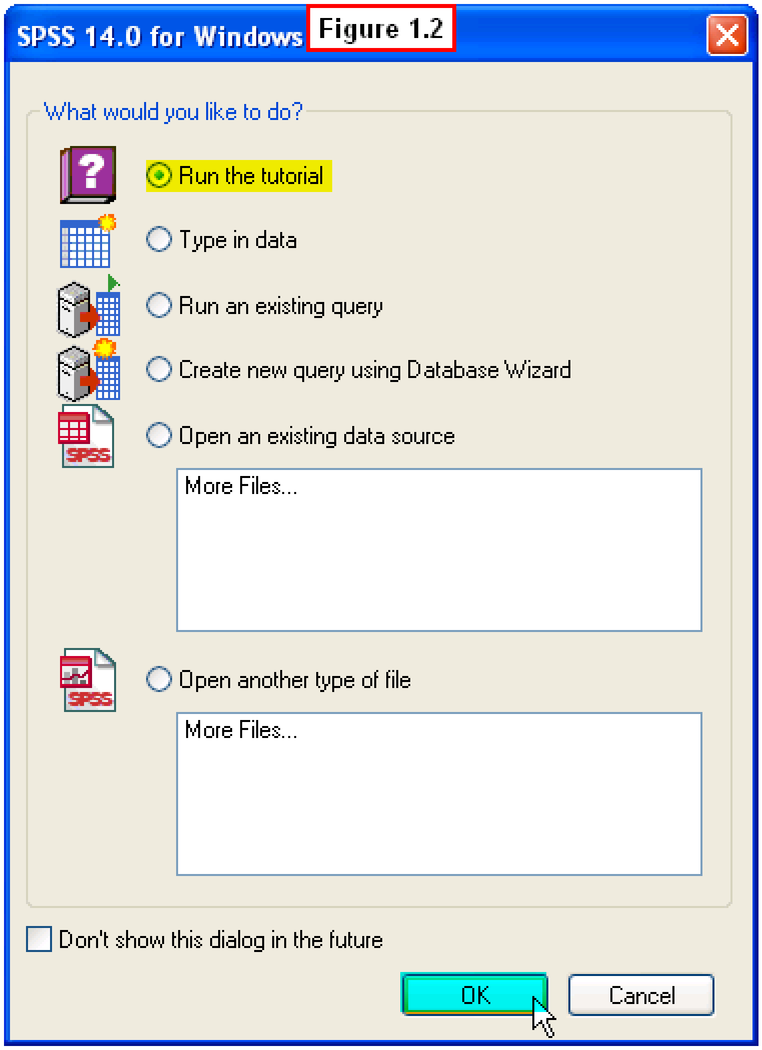
- Simultaneously press the **alt** and the **tab** buttons on the keyboard.
- When you press **alt-tab**, a pop-up will display all running programs/windows.
- While still depressing the **alt** key, tap the **tab** key to move between programs.

This will facilitate your moving back and forth between the browser window displaying this book and the window in which SPSS is open.

To open SPSS from your computer, click (in sequence) **Start, Programs, SPSS for Windows**, **SPSS for Windows** (Figure 1.1).



At startup a pop-up dialogue box appears asking "What would you like to do?" Select **Run the tutorial**, then click **OK** (Figure 1.2).



This start up screen will then close and the **SPSS Tutorial** window will open displaying the table of contents. To view a tutorial:

- Click on an item (e.g., **Introduction**), then click on a submenu item (e.g., **Starting SPSS**).
- A viewer window opens providing a tutorial on this topic.
- Click the right/left arrow buttons in the lower right of this screen to navigate through the topic.
- Click the **x** in the upper right corner to close the tutorial window.

1.5bThe Show Me Video! Feature of this Book

As an additional learning aid, throughout this text we include a feature called Show Me Video!

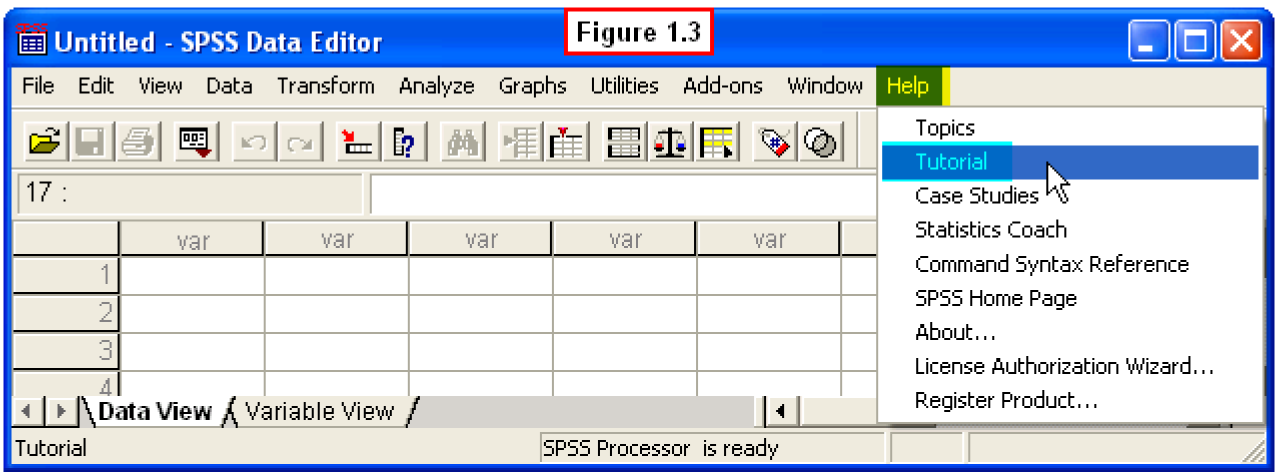
Throughout the book we provide short video clips of some of the SPSS procedures described in each chapter. After introducing a procedure and showing still images, we present a video that shows this procedure actually being performed in SPSS.

Below is the first of these videos, which shows the steps we just described above for running the SPSS tutorial.

Show Me Video!

1.5c Accessing the SPSS Tutorial in the Data Editor Window

When you close the tutorial, window you will return to the main window of SPSS (called the **Data Editor** window - we'll have more to say about this window in Chapter 2). You can also run the tutorial at any time by clicking on **Help** from the drop-down menu at the top of the window, then selecting **Tutorial** (Figure 1.3).



Show Me Video!

Not everything in the tutorial is relevant to the beginner, but we suggest that you run through the basics of SPSS explained in this tutorial. Also, keep in mind that there are many other ways to obtain tutorial help from SPSS other than by running the tutorial.

That is, **SPSS Help** can be accessed from just about anywhere within the program. In the **Appendix** we have included several of the screenshots and videos illustrating ways of obtaining help from various points within SPSS. We encourage you to review the Appendix so that you will be familiar with these other ways of obtaining help later on when we are performing various analyses in SPSS.

Now that you have had your first contact with SPSS, let's move on to a close encounter in Chapter 2, where you will learn how to create a simple data file in SPSS. We will then run a simple statistical procedure on this data and briefly discuss the resulting output.

This should familiarize you with the basic features of SPSS before we begin work on the EZDATA project mentioned earlier.

1.6 Chapter Video Review

Review Me! video is another learning feature of this text. At the end of every chapter, we will provide a **Review Video** which combines all the individual videos in the chapter into one video. This will be a good way of summarizing procedures and reinforcing your learning.

Review Me!

Chapter 2

Close Encounter:
The Data Editor, Syntax Editor & Output Viewer Windows

2.1 Introduction to SPSS

The capability of SPSS is truly astounding. The package enables you to obtain statistics ranging from simple descriptive numbers to complex analyses of multivariate matrices. You can plot the data in histograms, scatterplots, and other ways. You can combine files, split files, and sort files. You can modify existing variables and create new ones. In short, you can do just about anything you'd ever want with a set of data using this software package.

A number of specific SPSS procedures are presented in the chapters that follow. Most of these procedures are relevant to the kinds of statistical analyses covered in an introductory level statistics or research methods course typically found in the social and health sciences, natural sciences, or business.

Yet, we will touch on just a fraction of the many things that SPSS can do. Our aim is to help you become familiar with SPSS, and we hope that this introduction will both reinforce your understanding of statistics and lead you to see what a powerful tool SPSS is, how it can actually help you better understand your data, how it can enable you to test hypotheses that were once too difficult to consider, and how it can save you incredible amounts of time as well as reduce the likelihood of making errors in data analyses.

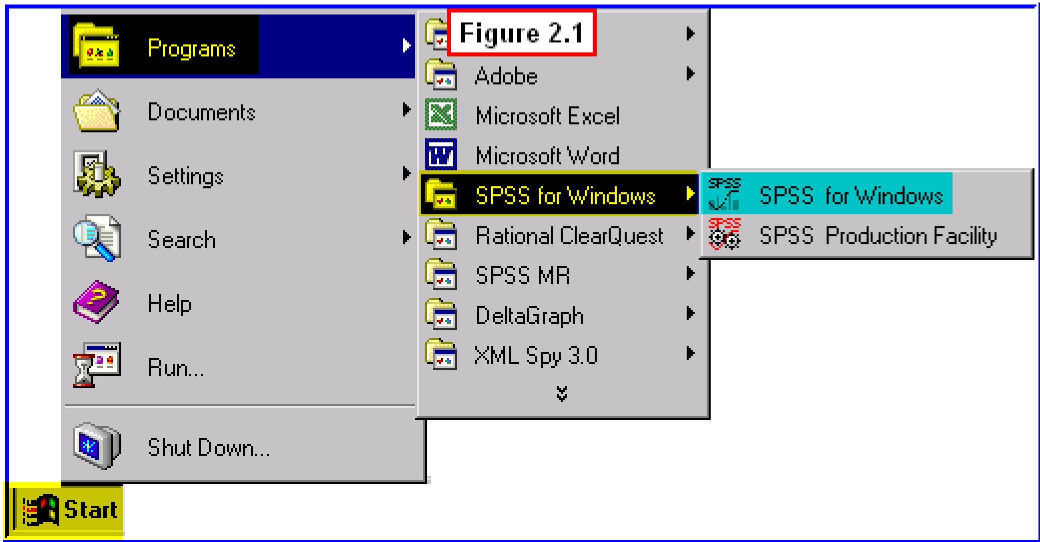
In this chapter we discuss the ways to open SPSS and we introduce the three main windows of SPSS. We show how to create a data file and generate an output file. We also discuss how to name and save the different types of files created in the three main SPSS windows.

2.2 The First Step: Opening SPSS

There are two methods of accessing SPSS, depending on whether you have the program installed on your hard drive or need to access it from a network.

2.2a Opening SPSS on a PC

If SPSS is installed on your hard drive, click **Start, Programs, SPSS for Windows, SPSS for Windows** (Figure 2.1).

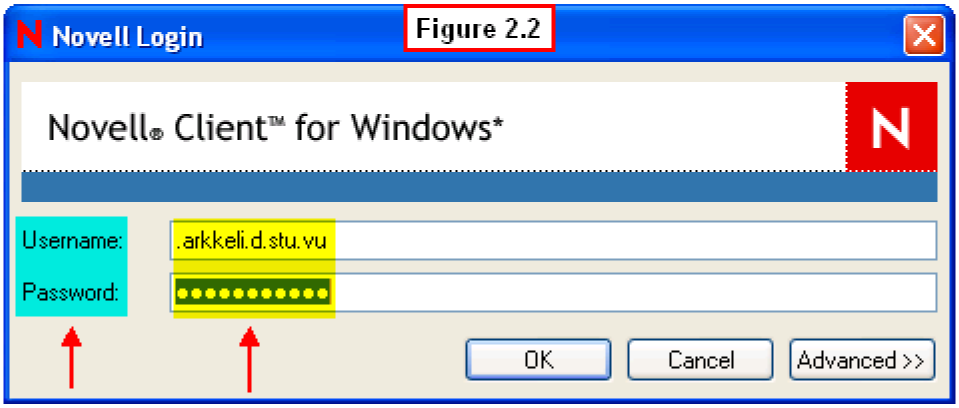


Recall from Chapter 1 that when you open SPSS, a start up screen appears asking "**What do you want to do?**" In Chapter 1 we answered that we wanted to run the tutorial. We will return to how to answer this question in the present chapter, but first we want to discuss how to access SPSS on a network. So if you have just loaded SPSS from your PC, just leave the start up screen open for now.

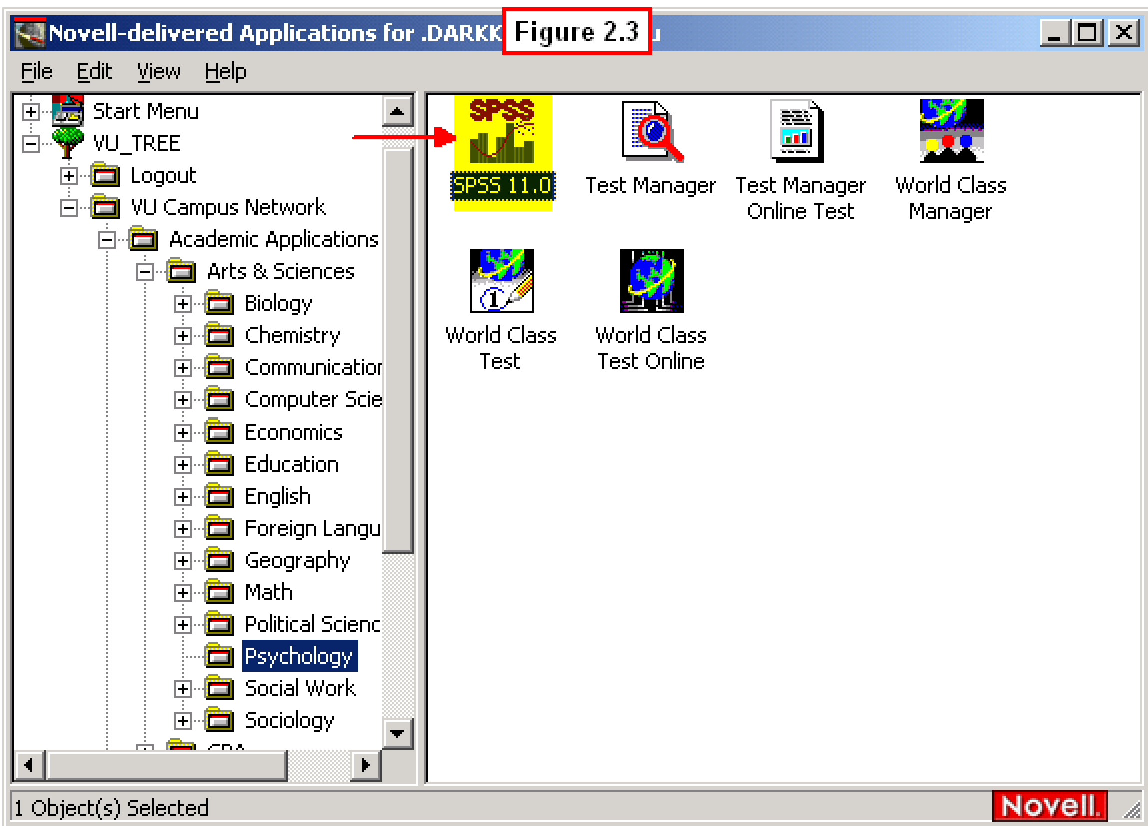
2.2b Logging in to a Local Area Network (LAN)

Different computers may have different operating systems, so you will have to learn the procedures specific to your computer system. If you do not already know how to log on to your computer system (or you do not know your username and password), we suggest that you learn how to do this before reading further.

In other words, you need to be able to log on to your computer network in order to access the SPSS software. Typically, your computer center or your instructor will be able to provide you with these procedures. As an example, below we illustrate how one would log on to the network and access SPSS for Windows at our own university. To log on to our campus network, users must type a Username and Password in the **Novell Login** window that appears when a computer connected to our LAN is booted up (Figure 2.2).



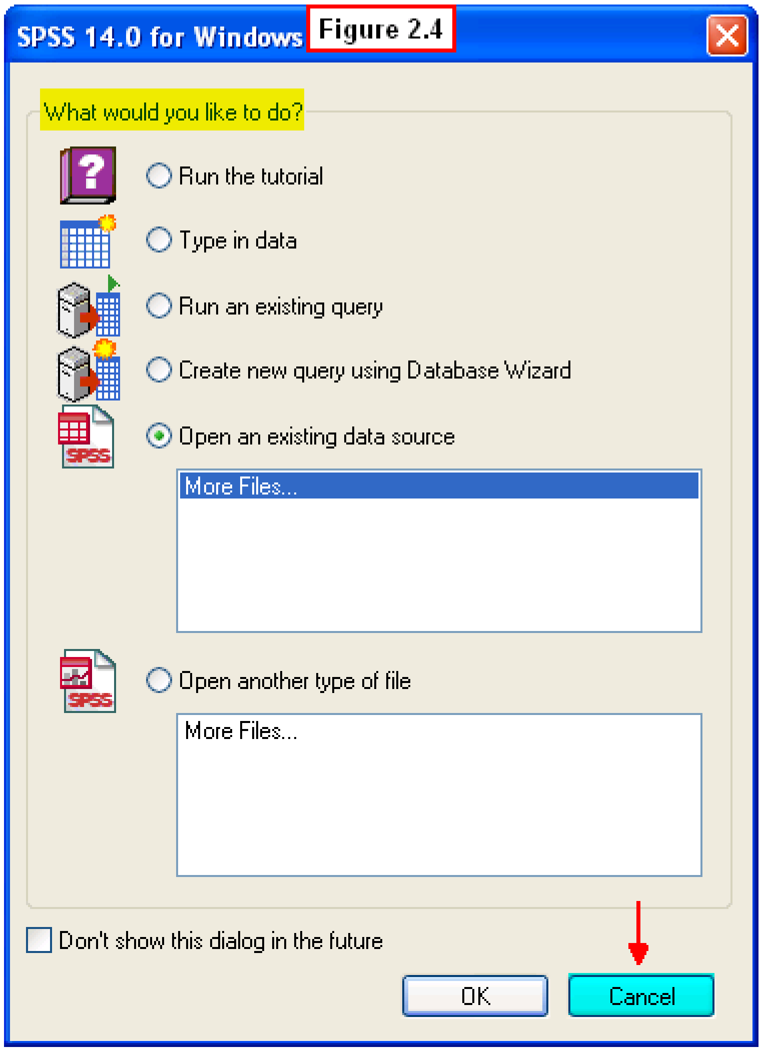
After entering the username and password, a **Novell-delivered Applications** window will pop up with an interface in the left panel similar to Windows Explorer tree menu (Figure 2.3).



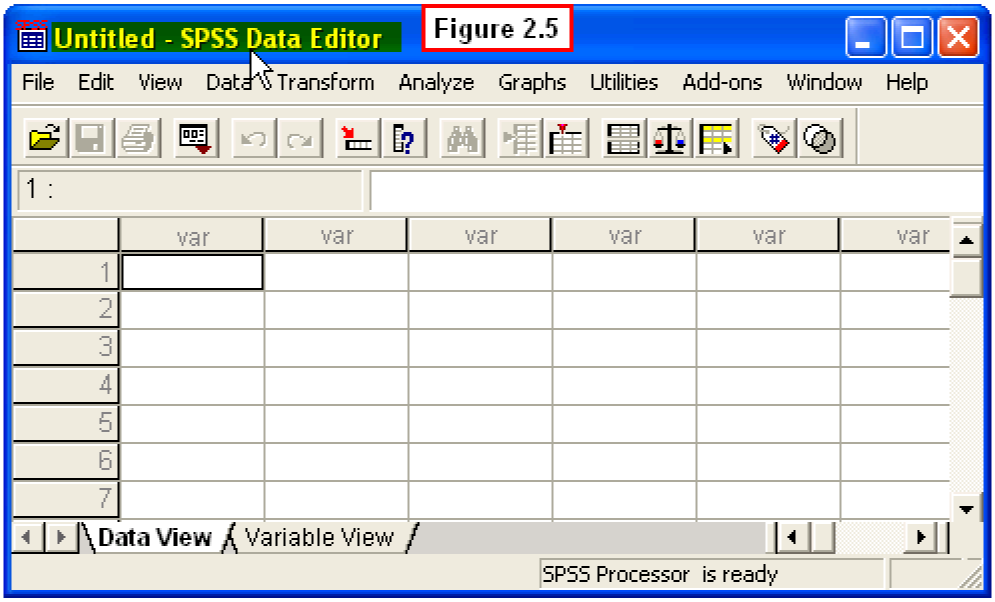
To access SPSS, the user clicks on the plus signs in front of various folders to navigate to the folder containing SPSS. At our institution, the user would click on the folders in the left panel of Figure 2.3 in this order: **VU Campus Network, Academic Applications, Arts & Sciences, Psychology**. After clicking the last folder, the **SPSS** icon appears in the right panel of the window. Double-clicking on this icon will start the SPSS program.

2.2 c SPSS start up screen

Recall that when you open SPSS, a dialog box appears with the question, **What would you like to do?** (Figure 2.4). This window allows the user to choose from a number of quick-start options, such as loading an existing data file or opening a recently-used file.



In Chapter 3 we will open an existing data file (the one we will create in this chapter). But since we don't yet have an existing file yet, just click **Cancel** in the lower right of this start up dialog box. The box will close, leaving a blank **Untitled - SPSS Data Editor** window open (Figure 2.5).l



[Show Me Video!]

2.3 The Second Step: Creating and Saving a Data File in the Data Editor

Now that we have accessed SPSS for Windows, we need to have a data file to analyze. SPSS recognizes and is able to import files created in other applications (e.g., Microsoft Excel, Microsoft Word, and Windows Notepad). We will not describe importing files from other applications - instead, we want to get you started by creating a data file from scratch.

Doing this will familiarize you with the main components of SPSS for creating data files, analyzing the data and viewing the results of those analyses. These three components consist of three types of windows:

- **Data Editor Window**
- **Syntax Editor Window**
- **Output Viewer Window**

In this section, we will discuss the first two windows, which are used to create data files (we'll discuss the Output Viewer window later). By learning how to create data files from scratch in the **Data Editor** and **Syntax Editor** windows, you will also come to understand how data is organized into a file. Files that can be imported into SPSS from other applications would be imported into either the Data or Syntax editors anyway, so working directly with these editors to create a file will help you understand any files that you would import from other applications.

Data files are created in SPSS by directly typing data into one of these two editors. The **Data Editor** works similar to a spreadsheet application, while the **Syntax Editor** works like a basic word processing application.

In the following we will describe the process of creating a data file in each of these editors. However, note that in subsequent chapters, we will work only with the Data Editor, since this is the primary window users of SPSS employ for data analysis. Use of the Syntax Editor for data analysis requires somewhat more advanced skills, so we will only refer to it occasionally in subsequent chapters. There is value in learning about this editor and how files are created with it, however, so we will address that in this section.

2.3a Data File Organization

The **Data Editor Window** that appears by default when SPSS is opened looks similar to and works like other spreadsheet applications such as Microsoft Excel. Data files are created by entering data into the cells of the table. To do this, simply click on a cell and type the appropriate numbers representing scores on variables to be analyzed. We will illustrate this process in the following.

Let's assume that a statistics professor is interested in the number of psychology courses that students have taken prior to enrolling in his/her course. S/he is also interested in comparing the number of previous psychology courses taken by male and female students. Table 2.1 consists of data, or scores, for the number of psychology courses taken by ten students, five men and five women.

Table 2.1

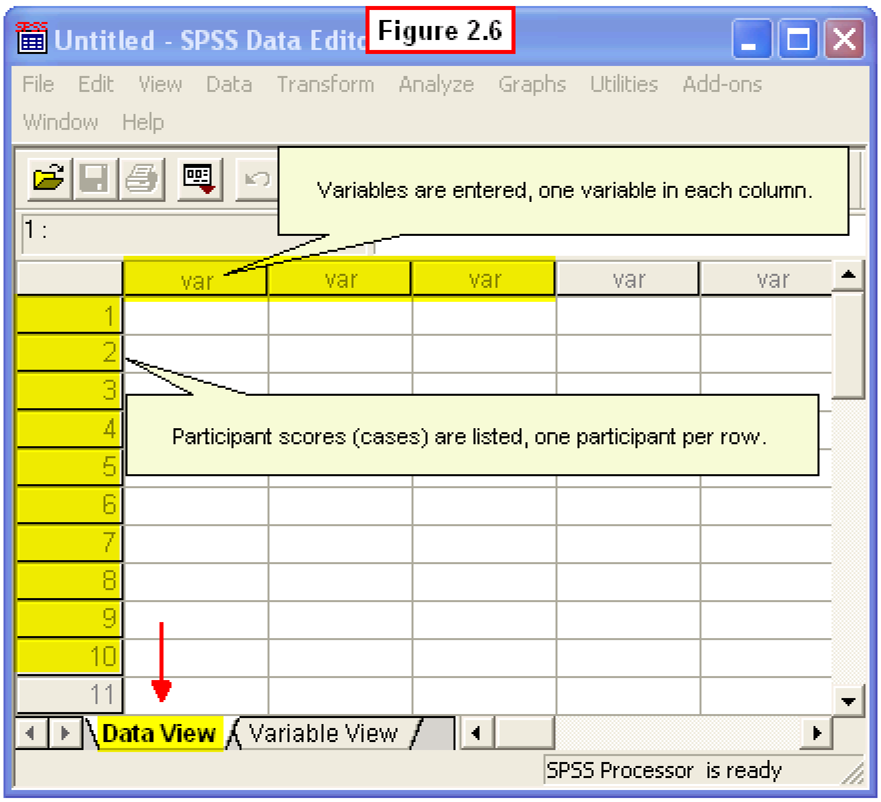
Student ID	Student Sex	No. of Courses
01	1	2
02	1	1
03	1	1
04	1	2
05	1	3
06	2	3
07	2	3
08	2	2
09	2	4
10	2	3

Note that there are three "variables" in this table: a student ID number (from 01 to 10), a code indicating the student's sex (where 1 = male and 2 = female), and the number of psychology courses each student has taken (ranging from 1 to 4).

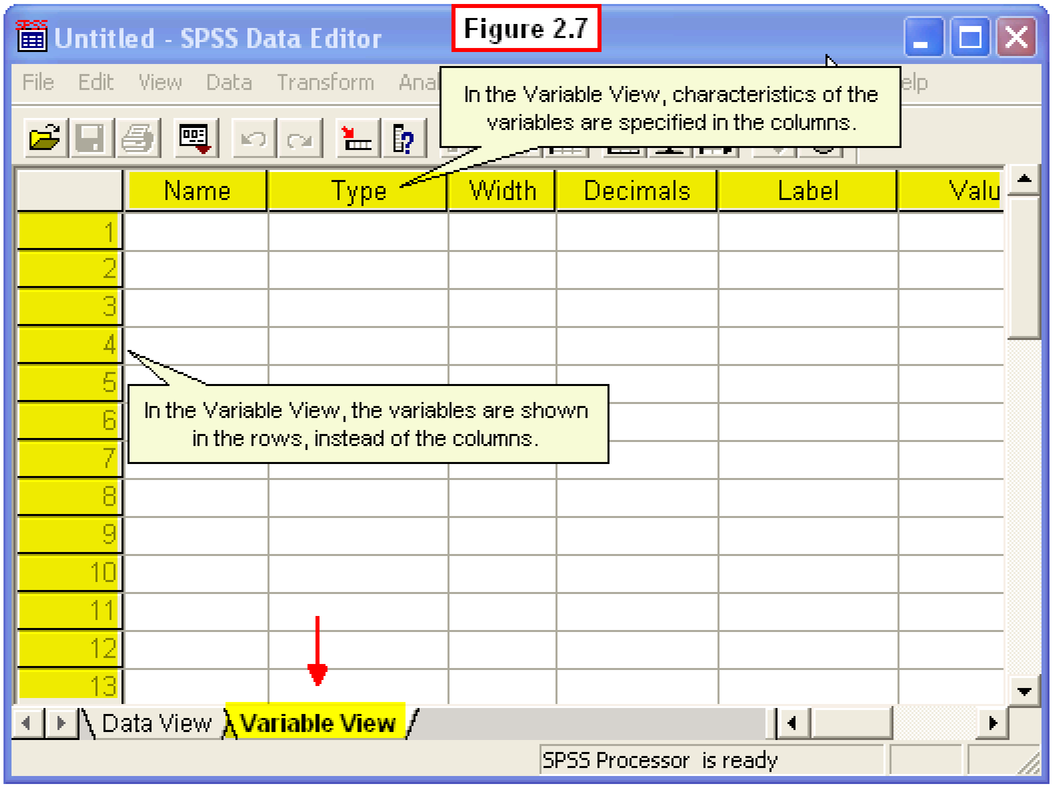
Thus, the data is organized with the three ***variables*** (ID, Sex, Courses) listed in ***columns***, and the ***scores*** for each of the ten students on each of these three variables entered in the ***rows***. You will see that the data needs to be organized in this same way in SPSS. If you haven't done so already, open SPSS now so you can follow along with this example.

2.3b Data Editor Rules

Let's create this file in the Data Editor window on your computer. As can be seen in Figure 2.6, SPSS expects you to list variables in the ***columns***, and individual scores from each participant in the ***rows*** of this spreadsheet. This is the same way that the data has been organized in Table 2.1, so this shouldn't be difficult to do.



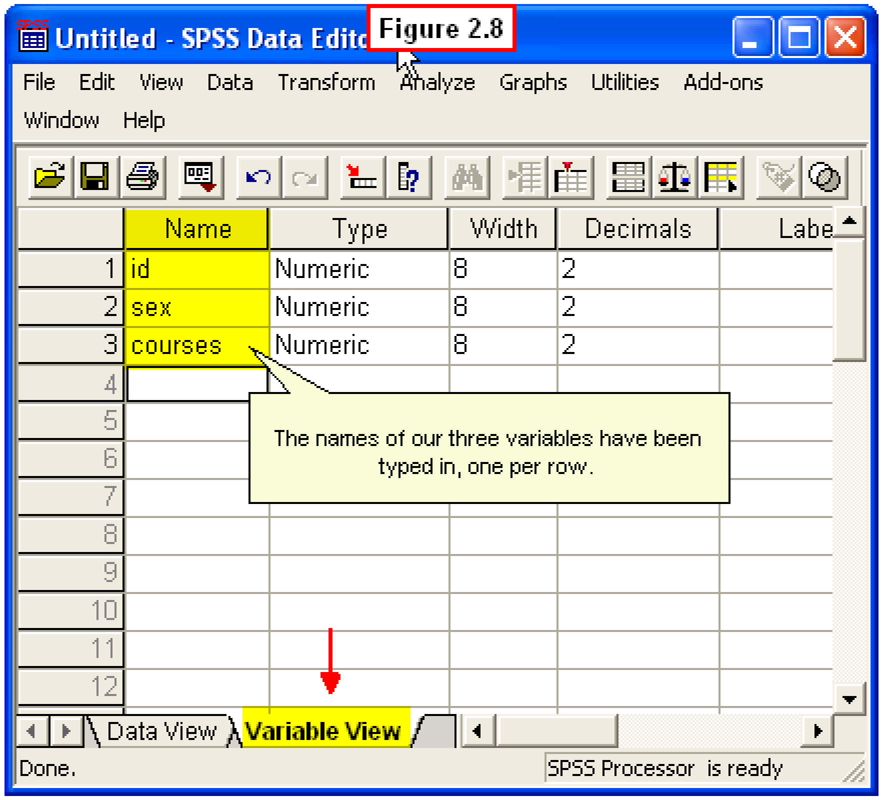
However, it's a good idea to name the variables in SPSS before entering the individual scores. More will be said about this in a later chapter, but for now, just note that the **Data View** tab is active at the bottom of this SPSS Data Editor window. To name the variables, we need to activate the **Variable View** tab. To do this, simply click on that tab. Figure 2.7 Shows the Data Editor window with the **Variable View** tab active.



It is important to note that in this view, the variables are listed in the ***rows*** (as opposed to being shown in the columns when the Data View tab is active). In the Variable View, instead of listing the variables in columns, ***characteristics*** of the variables are indicated in these columns.

Again, we will return to a more complete discussion of this view later, but for now, our main interest is in the column titled **Name.** This is where we will type the names of our three variables by clicking on the appropriate cell in the first column of each row.

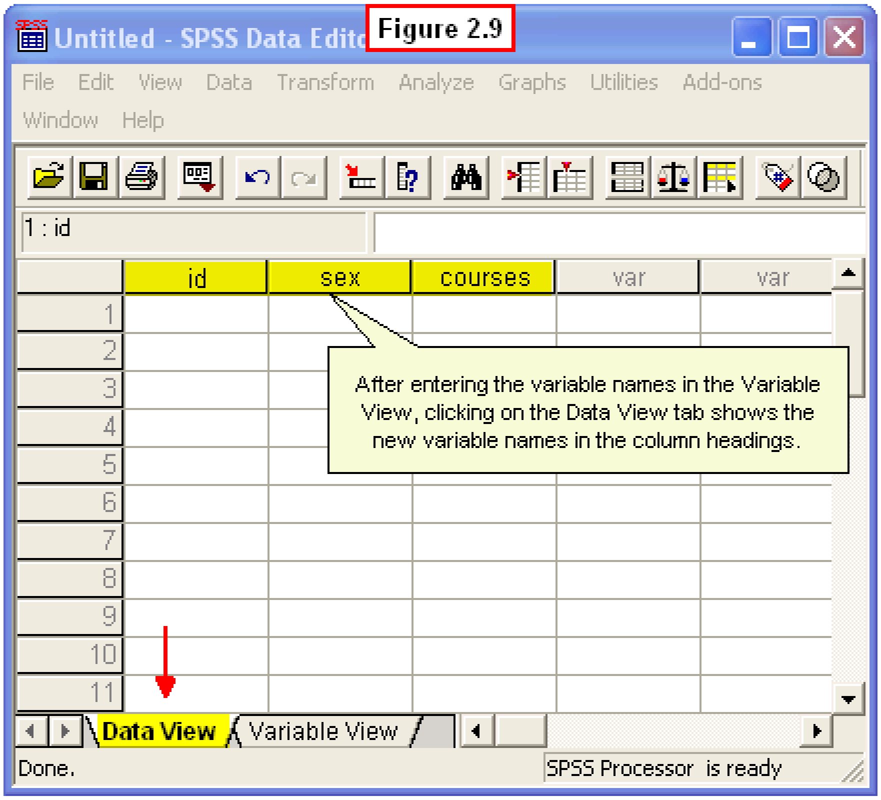
Figure 2.8 shows the variable names, **ID, Sex** and **Courses**, typed in the name column of the first three rows.



SPSS has rules for these variable names (e.g., there can be no spaces, the variable name must begin with a letter, and the maximum length is 8 characters). Upper or lower case may be used – SPSS doesn't care which you use, so we have used lower case.

Note that once a name has been typed in, SPSS default options appear in the next three columns. Don't worry about these for now. The most important one to note is that all three variables are numeric (as opposed to letters), which is correct.

Type these three variable names in your Data Editor window now. When you have finished, click the **Data View** tab at the bottom of the window. When you do this, you will see that the variable names you just typed in now appear as headings in the first three columns of the Data Editor window (Figure 2.9).



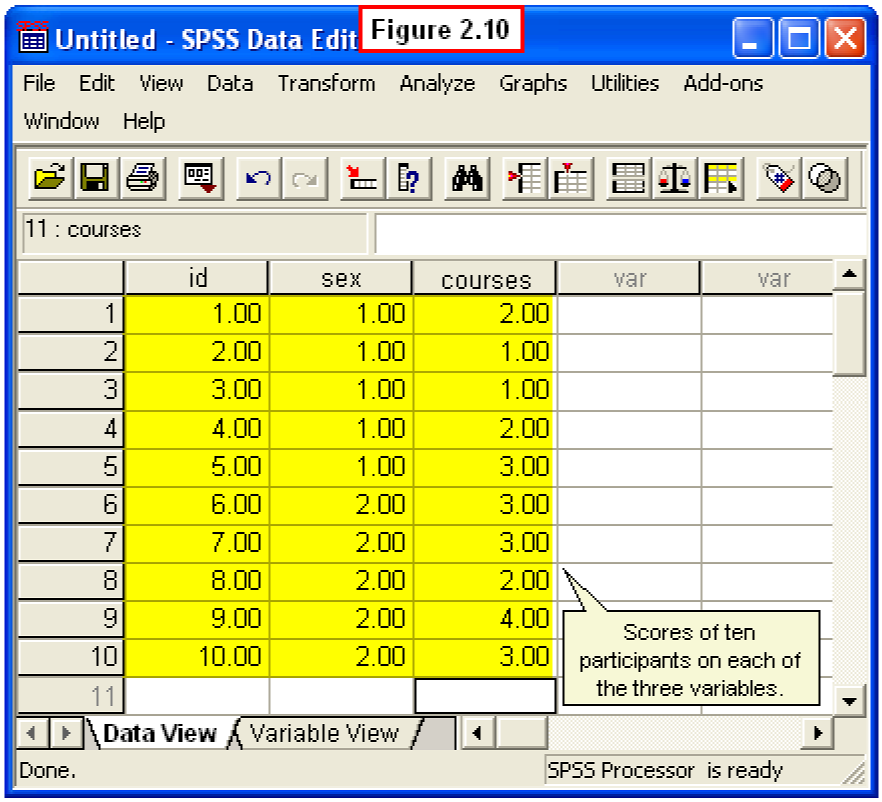
[Show Me Video!]

2.3c Entering Data in the Data Editor

Now we are ready to enter the data, or scores, for each student in the rows of this spreadsheet. To do this, simply type the scores shown in Table 2.1 into the appropriate cells. To begin, simply click the upper-most cell on the left and

- Enter the data in the first column
  - Type a 1 for the first **Student ID**; press enter.
  - Type a 2 for the second ID, press enter.
  - Continue typing the remaining ID numbers in this column.
- Enter the data in the second column
  - Click on the first cell of the column titled **Sex**; type a 1 for the first student's **Sex**.
  - Press enter and continue typing the remaining codes for student sex.
- Enter the data in the third column
  - Click on the first cell of the column titled **Courses**; type a 2 the number of courses.
  - Press enter and continue typing the remaining number of courses taken.

Alternatively, you could enter your data across each row rather than down the columns. That is, after typing the first student ID, you could use your right-arrow key to stay on that row and enter that student's sex; press the right-arrow again and type the number of courses for the first student. Then click on the first cell of the second row and enter the data across that row for the second student in the same way. When you have finished entering your data from Table 2.1, your Data Editor window should look like the one shown in Figure 2.10.



[Show Me Video!]

While data entry can be tedious, it is a critical first step in analyzing the results of research, so it is ***important to be careful to avoid text entry errors***. If you are not careful in entering data, the results of your analyses will be meaningless, so it is important to be careful and accurate in entering data. Further, after you have entered data, you should always double-check your file for accuracy. We cannot overemphasize the importance of accurate data entry. This is what the phrase, “garbage in, garbage out" is all about.

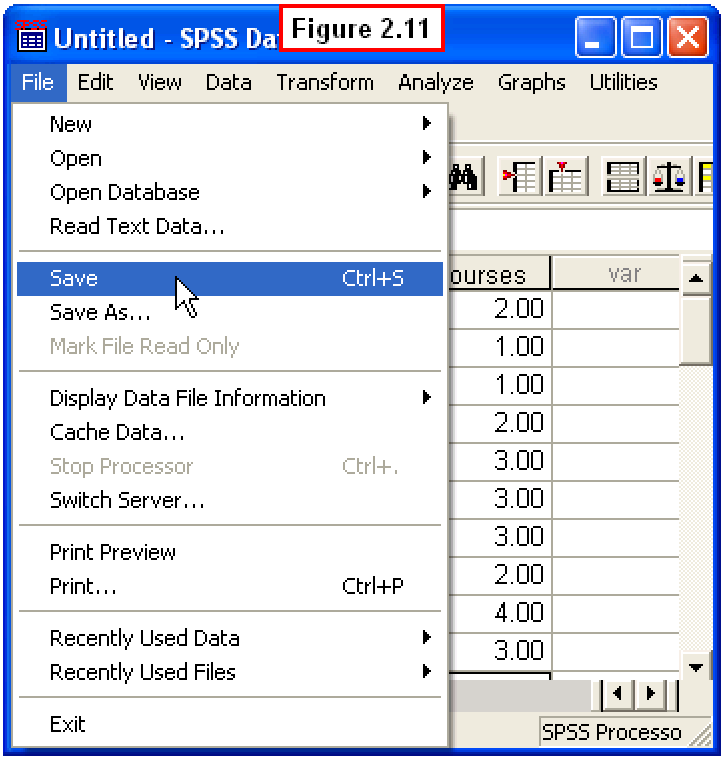
2.3d Saving files in the data editor: The .*SAV* File extension

When you name and save files created in the Data Editor to a hard drive or disk, SPSS by default adds the three letters, **.SAV**, (called the ***file extension***) to these file names. That is, just as Microsoft Excel adds the file extension, **.xls**, to file names created and saved from Excel, SPSS adds the extension, **.SAV** to files created and saved in the SPSS Data Editor.

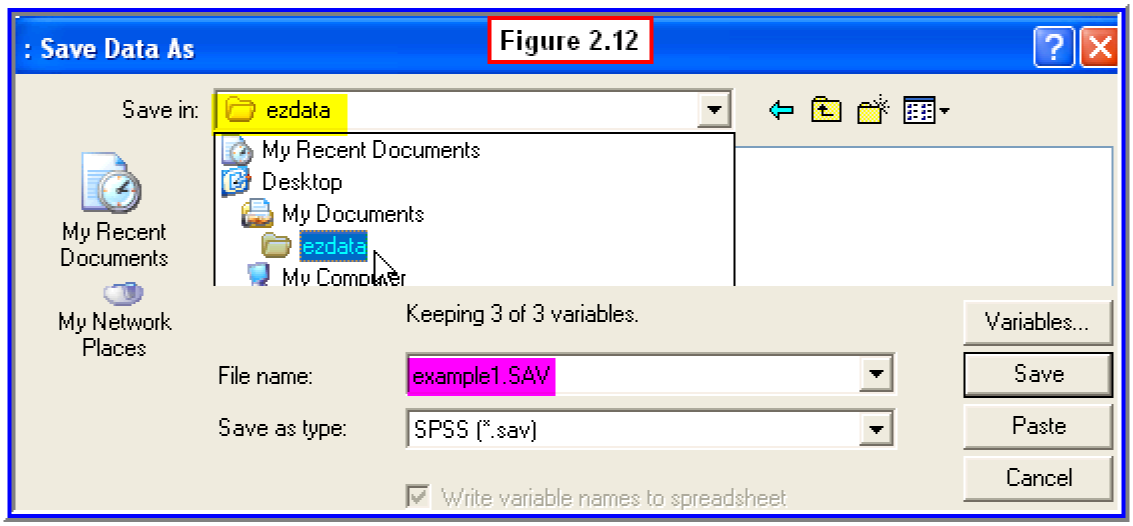
This is how your computer knows which program or editor to use when opening a given file. Each application (e.g., Excel, Notepad, SPSS Data Editor) has a unique file extension which causes the computer to use the appropriate program to open the file.

We recommend that you create a folder in which to save all of the examples and exercises you complete in this text. This folder can be on a floppy disk, your computer hard drive, a network drive, or some other media. This way you can always go back to look over examples you have completed.

To save your Data Editor file, follow the same procedure you would in saving any other file in another software application. From the drop-down menu at the top of your Data Editor window, select **File, Save** (Figure 2.11).



Then navigate to the folder you created for saving files. For this example, we recommend typing **example1.SAV** in the filename box of your **Save Data As** box (Figure 2.12).



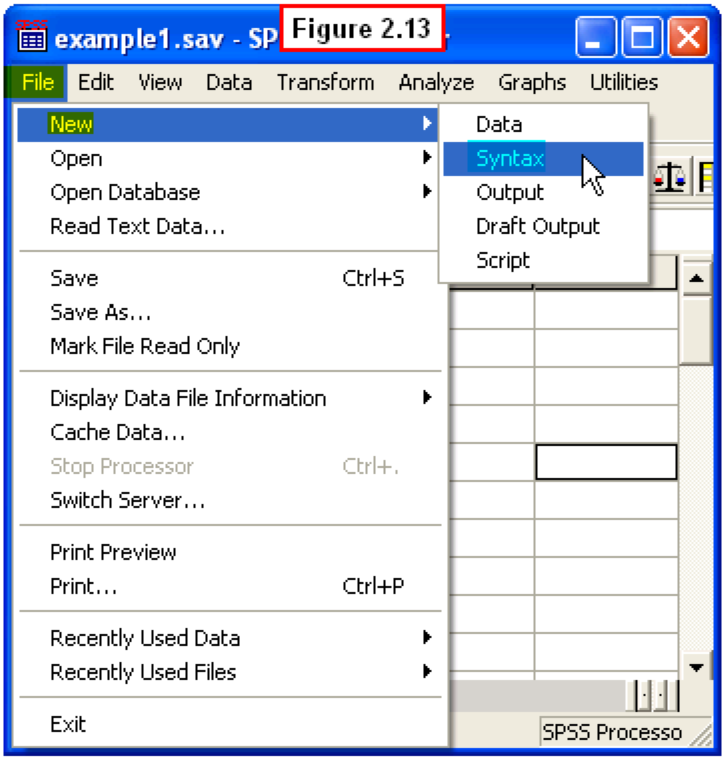
Show Me Video!

2.4 Creating files in the Syntax Editor Window

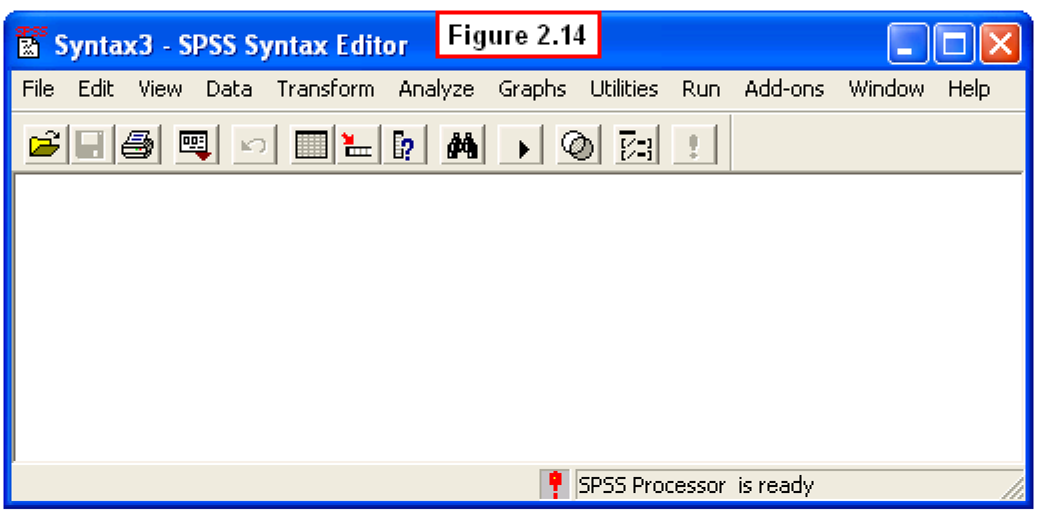
An alternate way to create data files is to use the SPSS **Syntax Editor Window.** The Syntax Editor works much like a word processor, such as Windows Notepad or a scaled-down version of Microsoft Word. That is, you can just type the data into this window as you would in creating any text file.

2.4a Accessing the Syntax Editor

Select **File, New, Syntax** from the drop-down menu at the top of the Data Editor window (Figure 2.13).



A new window will open (Figure 2.14).

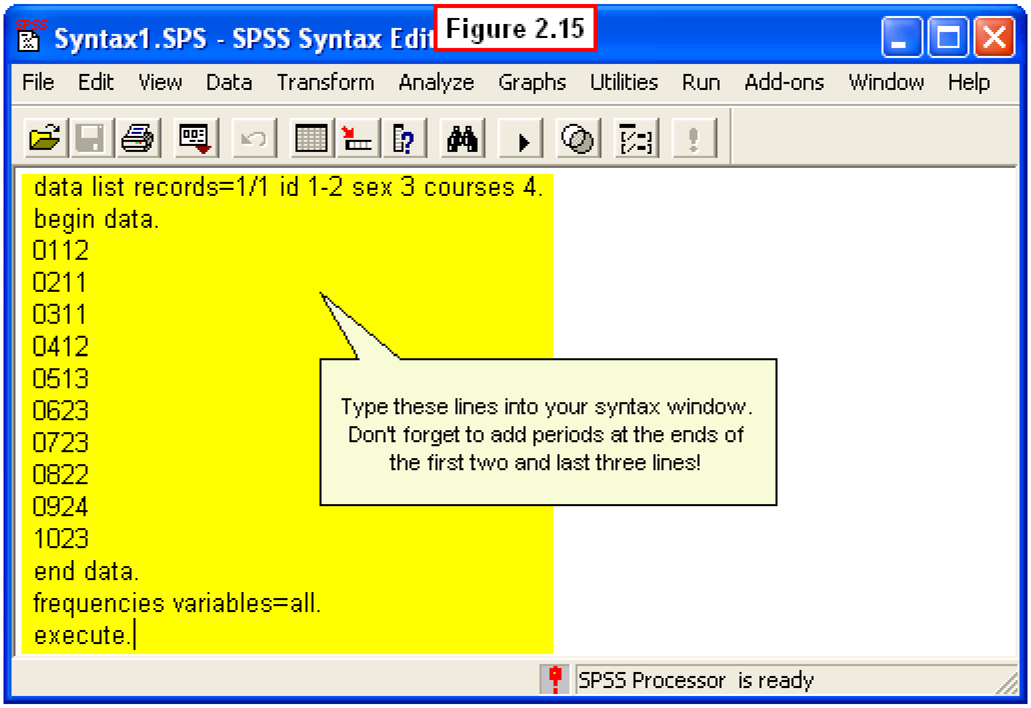


[Show Me Video!]

Note that in addition to typing text/data directly into the Syntax Editor window, you can "cut-and-paste" data into this editor. That is, just as you can cut-and-paste text from one Word file into another Word file, you can also cut-and-paste data into the Syntax Editor. Thus, data files that were created in a text editor such as Windows Notepad (called ASCI II files) can be cut-and-pasted into this Syntax Editor. However, as mentioned, we will create this Syntax Editor file from scratch by typing the text and data directly into this editor. Let's use the same example we just used to create this same data file in the Syntax Editor.

2.4b Entering Commands & Data in the Syntax File

We will be working with the Data Editor window in this book, so we will not discuss the Syntax Editor in detail here. For now, just type the file exactly as it appears in in Figure 2.15 into your blank Syntax Editor.



[Show Me Video!]

The first line of this file consists of syntax code, or SPSS commands, called the **DATA LIST** statement. This instructs SPSS how to read the data in the file. Note that the three variables (ID, Sex, and Courses) are defined on this line.

The actual lines of data (scores) are always sandwiched between a Begin Data and an End Data syntax commands. Note that the data are organized exactly like they are in Table 2.1 and Figure 2.10. That is, each line of data (or row) consists of each student's ID number, code for sex, and number of psychology courses.

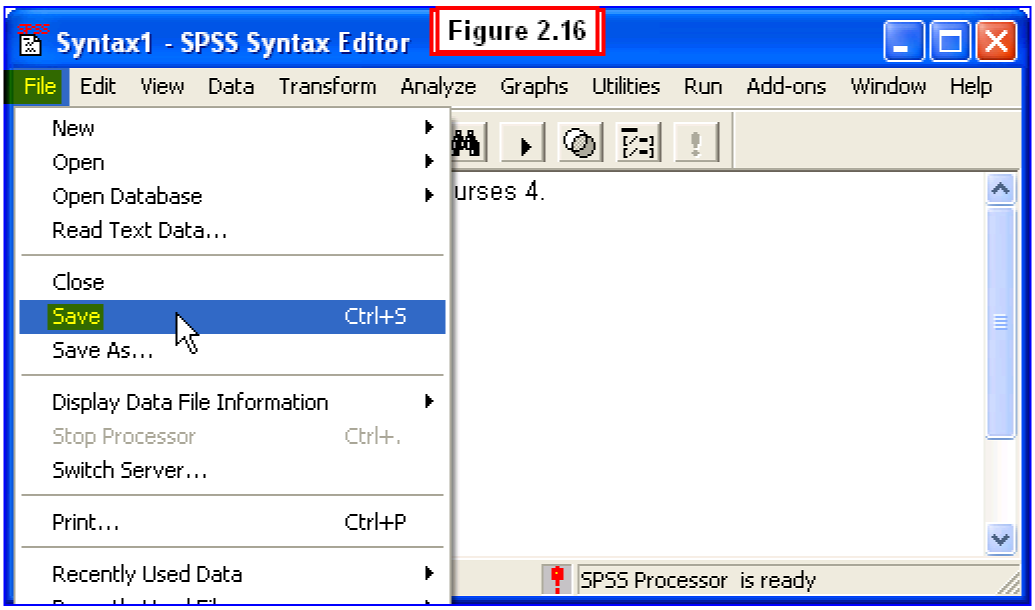
The last two lines instruct SPSS to generate **frequency tables** of each variable. Note that except for the lines listing the data, **the lines containing syntax code must all end with a period.** After you have created your file, you will be ready to save it as a syntax file.

2.4 c Saving Files in the Syntax Editor: The .*SPS* File extension

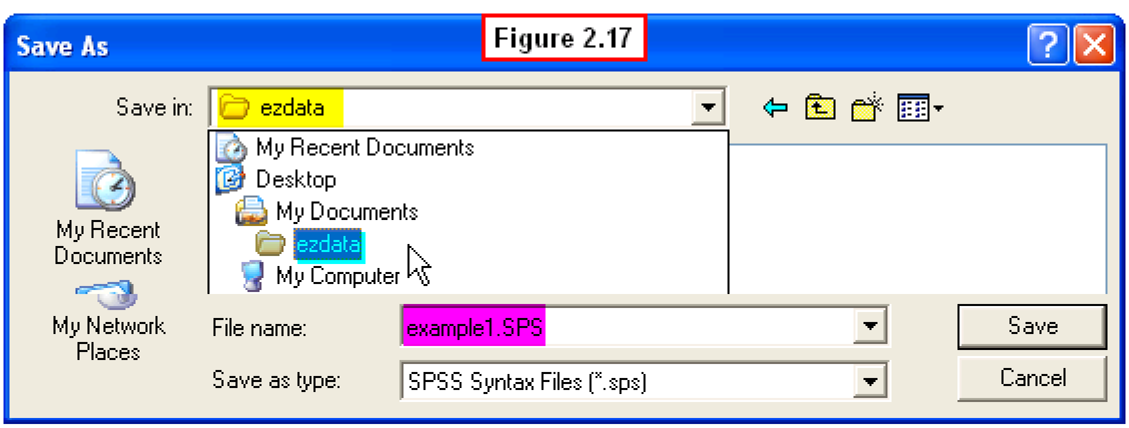
When you name and save files created in the Syntax Editor to a hard drive or disk, SPSS by default adds the file extension, ***.SPS***, to these file names. That is, just as Microsoft Word adds the file extension, **.doc** to filenames created and saved in Word, SPSS adds the extension ***.SPS*** to files created and saved in the Syntax Editor.

As mentioned in our discussion of ***.SAV*** files saved from the Data Editor, it is important that you learn this file extension so that you recognize that a file name with the extension, ***.SPS*** (e.g., Example1.SPS) was created in the SPSS Syntax Editor and will automatically be opened with the Syntax Editor. This is important so that you do not confuse these with ***.SAV*** files created in the Data Editor.

To save this Syntax Editor file, select **File, Save** from the drop-down menu at the top of the Syntax Editor window (Figure 2.16).



Then navigate to the folder where you are keeping files for this book, and type **Example1.SPS**in the variable name box of the **Save As** box (Figure 2.17).



[Show Me Video!]

Leave this Syntax Editor window open, as we will use it to demonstrate the third major component of SPSS in the last section: the **SPSS Output Viewer** window.

2.5 And Last, the SPSS Output Viewer Window

Once a syntax data file has been created consisting of a DATA LIST statement, lines of data, and an SPSS command statement specifying a procedure to execute, the user is ready to conduct the requested analysis.

2.5a Generating an output file

Let's go ahead and run the Frequencies procedure specified in this Syntax file so that we can generate an output file to view. We won't go into detail about this output now - we'll be doing plenty of that in later chapters! Our purpose here is to bring this example to completion and to show you a SPSS **Output Viewer Window**.

To generate this output, we need to instruct SPSS to run the procedure specified in the **Frequencies** syntax line at the end of the file. To do this, select **Run, All** from the drop-down menu at the top of this Syntax Editor Window (Figure 2.18).

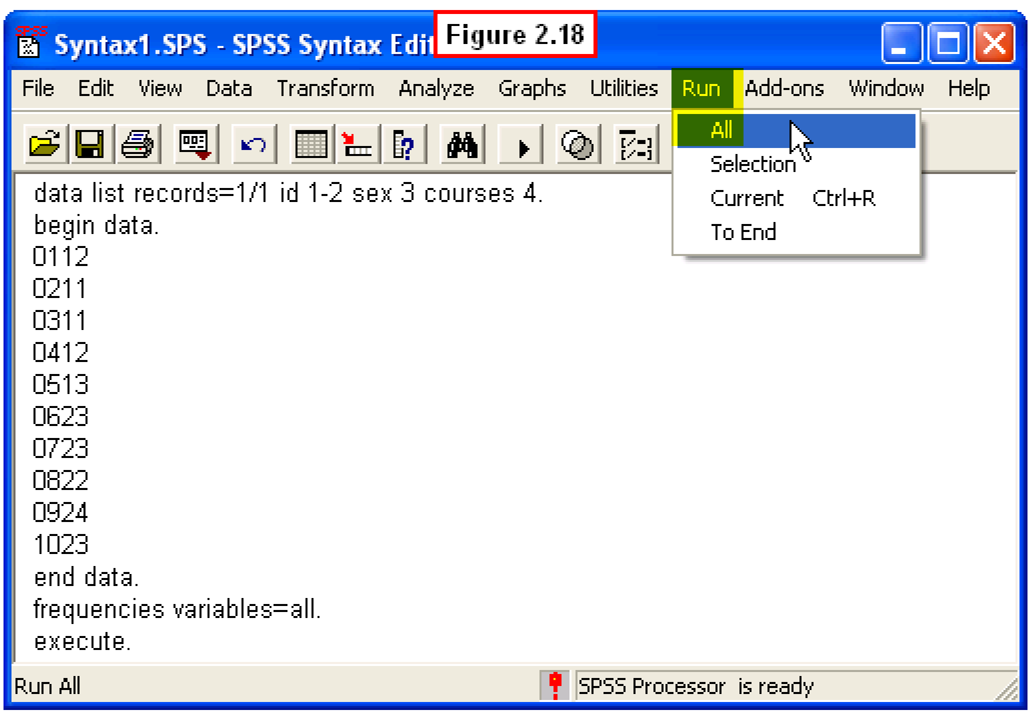
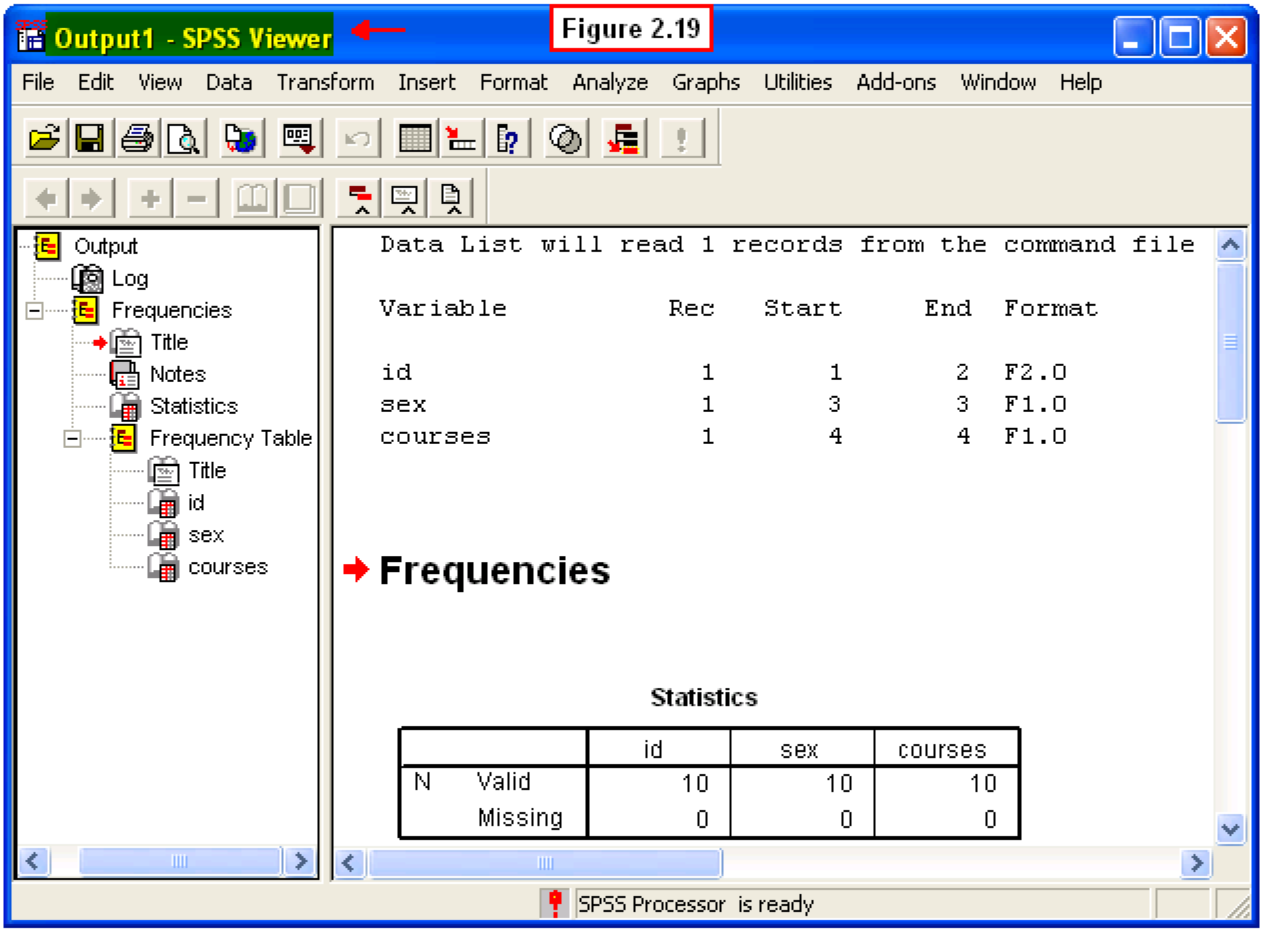


Figure 2.18

When you do this, the SPSS processor will run the procedure. Shortly a new window will appear presenting the results (**output**) of the **Frequencies** procedure.

The **Output Viewer** window shown in Figure 2.19 will now appear. This file shows the frequency tables we requested SPSS to generate. We won't discuss this output – our goal here is to contrast the **Output Viewer** window and its files from files created in the **Data Editor** and **Syntax Editor** windows.



[Show Me Video!]

2.5b Naming and Saving files in the Output Viewer: The *.SPO* File extension

Just as SPSS uses unique file extensions for Data Editor files (***.SAV***) and Syntax Editor files (***.SPS***), when you save an output file in the **SPSS Viewer** window, SPSS adds the extension ***.SPO*** to the file so that you and SPSS can recognize the file as an output file.

We emphasize that it is important that you learn this file extension so that you recognize that a file name with the extension, ***.SPO*** was saved in the SPSS **Viewer** window. This is important so you do not confuse these with ***.SAV*** files created in the Data Editor or **.SPS** files created in the Syntax Editor.

To save this file, select **File, Save** from the drop-down menu at the top of this window (Figure 2.20).

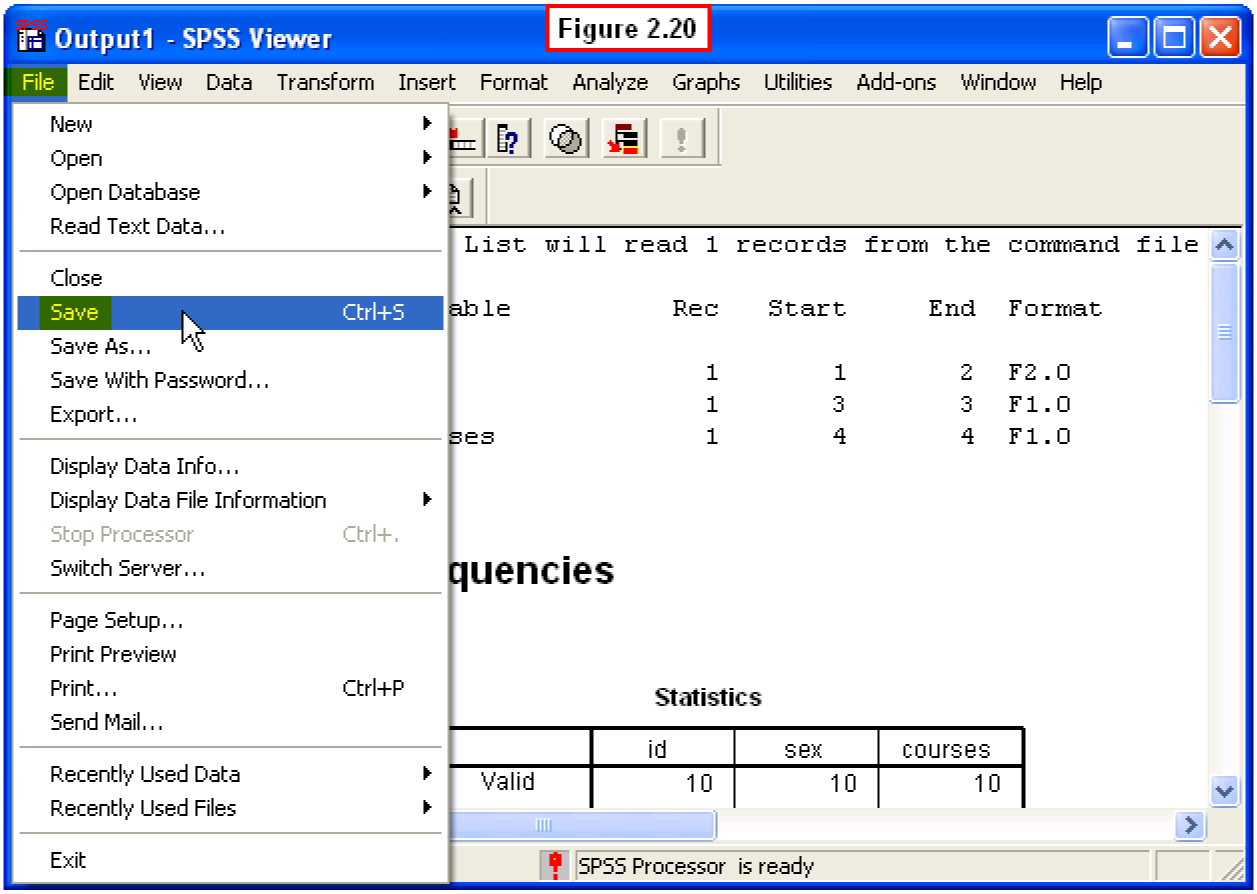


Figure 2.20

Navigate to the appropriate folder and type **example1.SPO** in the variable name box of the **Save As** box (Figure 2.21).

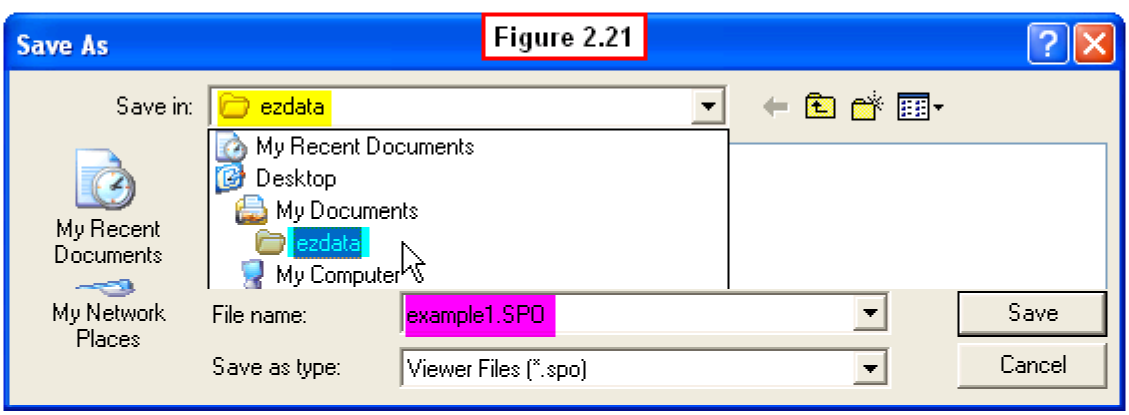


Figure 2.21

[Show Me Video!]

2.6 Summary of the three SPSS windows

This chapter has introduced the three major components of SPSS:

- **Data Editor** - a spreadsheet used to create data files and run analyses using menus
- **Syntax Editor** - a text editor used to create files and run analyses using syntax code
- **Output Viewer** - a window displaying the results of analyses performed by SPSS

As mentioned, we will use the **Data Editor** window in the remaining chapters of this book, and will refer only occasionally to the **Syntax Editor** window. Regardless of whether we are working with the Syntax Editor or the Data Editor, the results of our procedures will always be displayed in the **Output Viewer** window. In both cases, the results will be displayed in the Viewer window.

To summarize what we have covered in this chapter, you will progress through the following steps as you create and execute an SPSS program:

- **Access SPSS** for Windows from your PC hard drive or from your network (LAN)
- **Create a data file** in the SPSS Data Editor or the SPSS Syntax Editor
- **Save the data file** using the appropriate file extension (**.SAV** or **.SPS**)
- **Specify an analysis** to be run on the data file (e.g., **Frequencies**)
- **View the results** of an analysis in the SPSS Viewer Window
- **Save the output file** in this Viewer Window using the **.SPO** file extension.

Once again, although this may seem confusing at first, we assure you that all of this will soon become second nature to you once you have conducted a few analyses using SPSS! We conclude the present chapter by explaining some of the differences between working with files in the Data Editor versus the Syntax Editor.

2.7 The SPSS Data and Syntax Editors: Point-and-Click vs. Syntax Code Methods

The Data Editor window allows the researcher to work with SPSS using a method commonly referred to as **point-and-click**, meaning that you use your computer's mouse to **point** the cursor at various buttons/icons, and then you select options from drop-down menus by **clicking** the appropriate options.

This use of SPSS for Windows works just like most other software programs that operate in Windows that you have undoubtedly used many times (e.g., Microsoft Word). It might not surprise you to learn that when you use the point-and-click method in SPSS, you are actually generating a sequence of syntax or command statements that will eventually be used to execute a program (or run the procedure) using the data entered through the Data Editor window.

Although you don't see them as part of the point-and-click method, as you will learn later, you are able to view these "behind the windows" syntax commands and save them in a Syntax Editor file. These commands are actually fairly intuitive (e.g., we saw earlier that the word, **Frequencies,** is the syntax command to generate frequencies tables).

Typing these syntax command codes directly into a Syntax Editor window is referred to as the **syntax code method**. The procedure produces the same analyses as those generated by the point-and-click method (without having to do all the pointing and clicking!). The user just needs to learn the relevant syntax command codes to type them into the Syntax Editor.

The point-and-click method is in some ways is the easiest way to use SPSS for Windows, though it can be tedious to point and click through the many steps in creating a file, modifying the file, selecting the type of analysis to conduct and specifying the various options available at each step of this process.

Further, there are some real advantages of the syntax code method, so to fully tap all of SPSS's capabilities, it's worth learning about SPSS syntax. For example, some procedures are simply not available using the point-and-click method, and can only be performed using the syntax code method. Thus, although we have chosen to introduce the point-and-click method to avoid overwhelming the beginner, the more advanced student is encouraged to learn about the syntax code method as well.

Just remember that which method you are using will determine which SPSS window you will employ:

- **Data Editor:** User selects options by pointing and clicking on menus/buttons
- **Syntax Editor:** User selects options by typing in appropriate syntax command codes

From here on, we will present only the **point-and-click method** in the **Data Editor** window. For example, in Chapter 3 we will show you how to use the point-and-click method to generate the frequency tables we produced using the syntax code method in the example in the present chapter.

2.8 Chapter Review Video

[Review Me!](#)

2.9 Try It! Exercises

1. Creating a data file in the Data Editor

Open SPSS and use the Data Editor to create the **example1.sav** file described in Section 2.3. After you have created the file:

- **Save** the file to a disk, your hard drive or a network drive - ***you will need this file in a later chapter!*** We recommend that you create a folder titled **ezdata** in which to store all files you generate as you follow along with examples and do the exercises. These can be helpful review.
- **Print** this file by selecting **File, Print** from the Data Editor menu to submit to your instructor.

2. Creating a data file in the Syntax Editor

Use the Syntax Editor to create the file **example1.sps** file described in Section 2.4. Leave this file open.

3. Creating an Output file

With the **example1.sps** file open in the Syntax Editor, generate an output file by selecting **Run, All** from the menu. SPSS will perform the analysis and display the results in an Output Viewer window. Your file should look like the one shown in Video 2.9.

- **Print** this file to submit to your instructor by selecting **File, Print** from the Output Viewer menu.

Chapter 3
Point-and Click:
Conducting Analyses in the Data Editor

3.1 Running the Frequencies Procedure in the Data Editor

In Chapter 2 we created a small data file consisting of the number of psychology courses taken by five male and five female statistics students. We will continue with this example in this chapter. Recall that in Chapter 2 we:

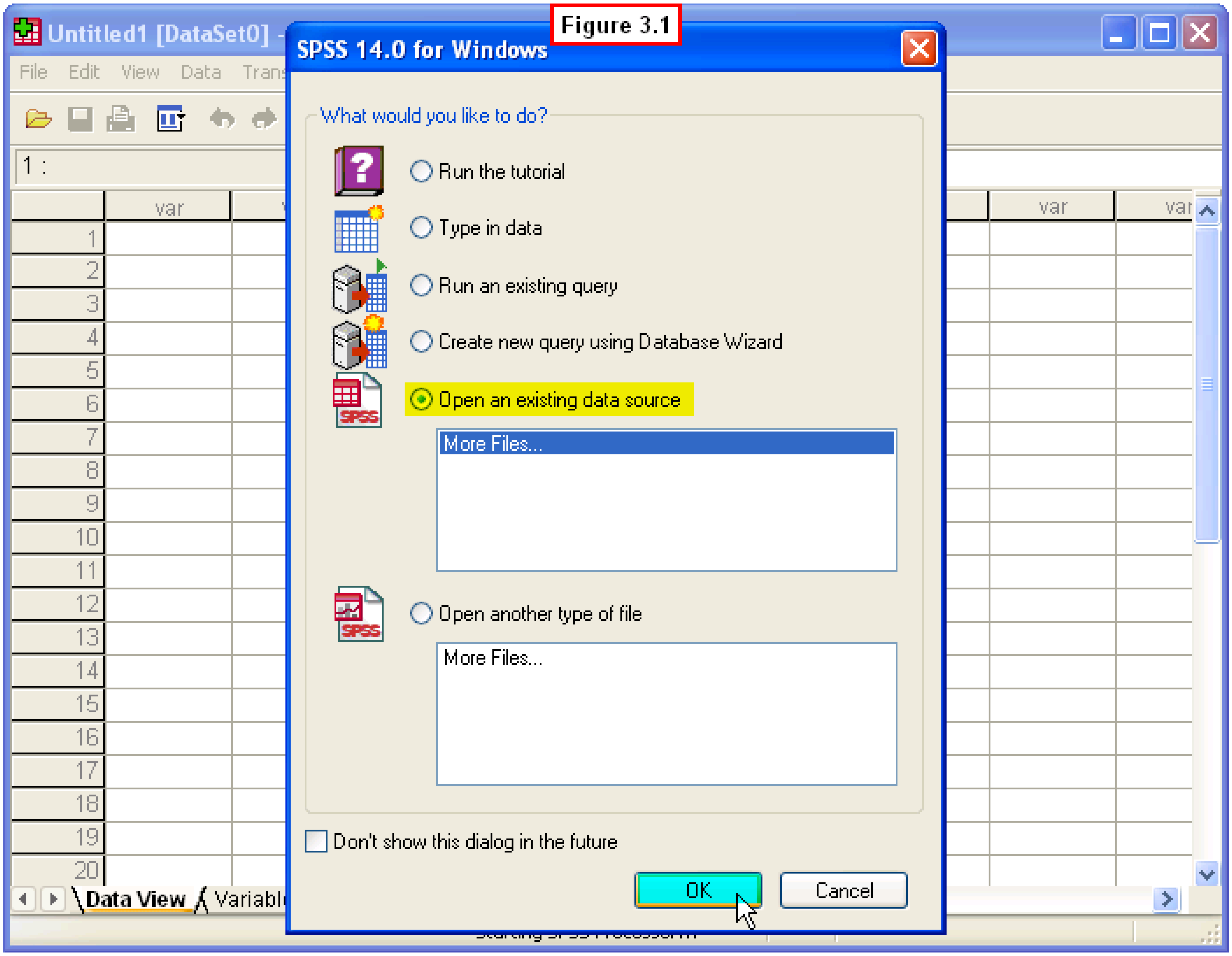
- created this data file in the **Data Editor** window (**example1.SAV**)
- created the same file in the **Syntax Editor** window (**example1.SPS**)
- generated frequency tables, displayed in the **Output Viewer** window (**example1.SPO**).

In Chapter 2, we generated the frequencies tables using the Syntax Editor. In this chapter we will conduct the same procedure in the Data Editor using the **point-and-click** method. We will use the **example1.SAV** file we created in Chapter 2 to do this. This method requires no knowledge of SPSS syntax code (recall that SPSS writes this code in the background as you point-and-click various selections in the Data Editor window).

3.1a Opening an existing data file

To follow along with the example in this chapter:

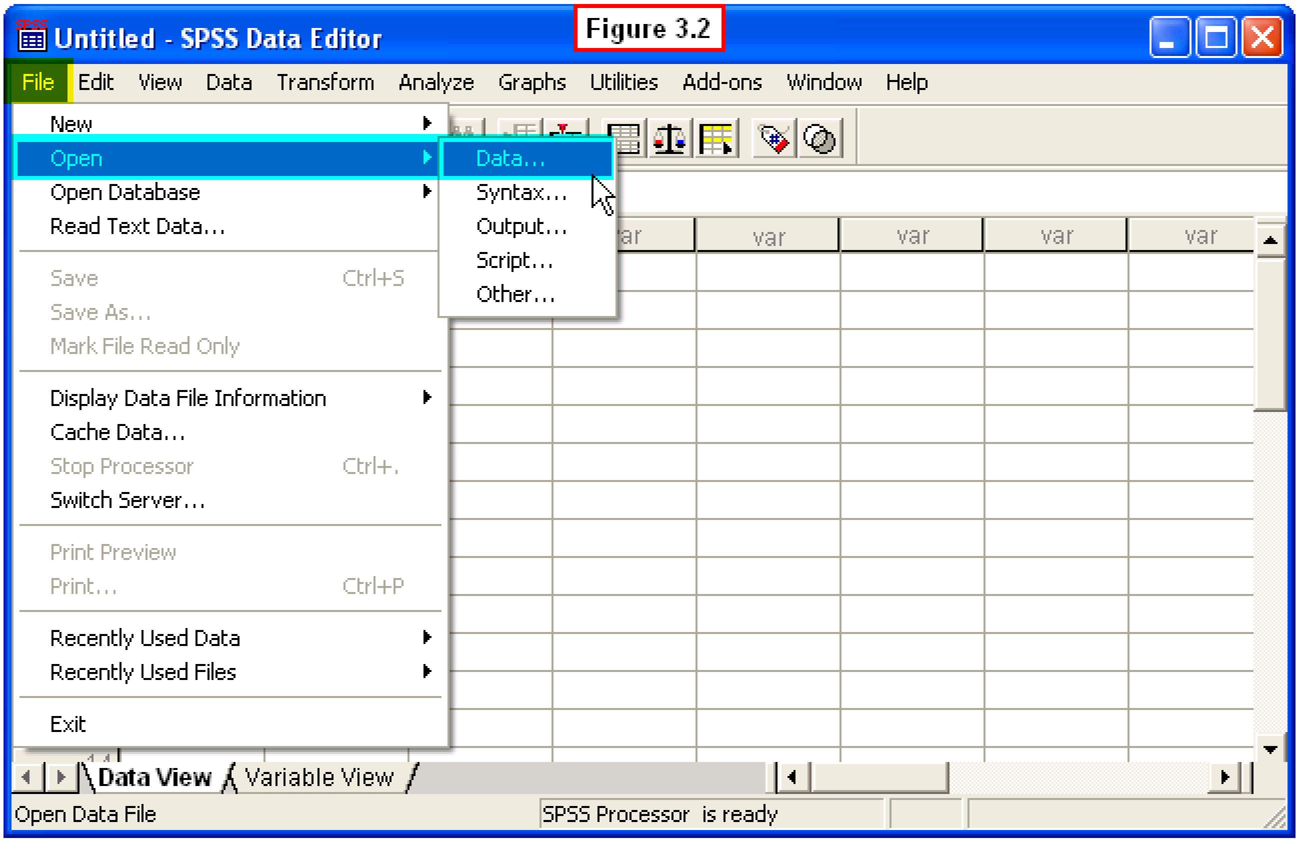
- open SPSS
- reply to the "What do you want to do?" question by selecting
- **Open an existing data source** (Figure 3.1)



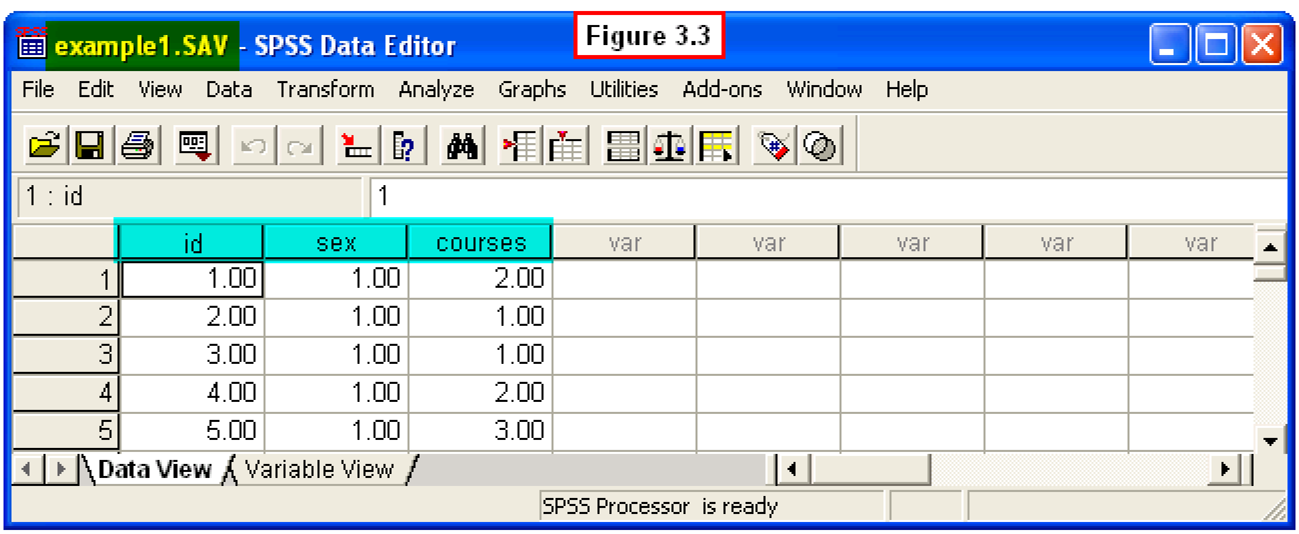
Then navigate to the folder containing your **example1.sav** file and select that file to open.

Alternatively, if you clicked **Cancel** in response to the start up dialog box, you can still open an existing file from the drop-down menu in the Data Editor by doing the following:

- Select **File, Open, Data…** from the drop-down menu (Figure 3.2)
- Navigate to where you saved the **example1.sav** and select that file to open.



Your **example1.SAV** data file should now appear in the Data Editor (Figure 3.3).

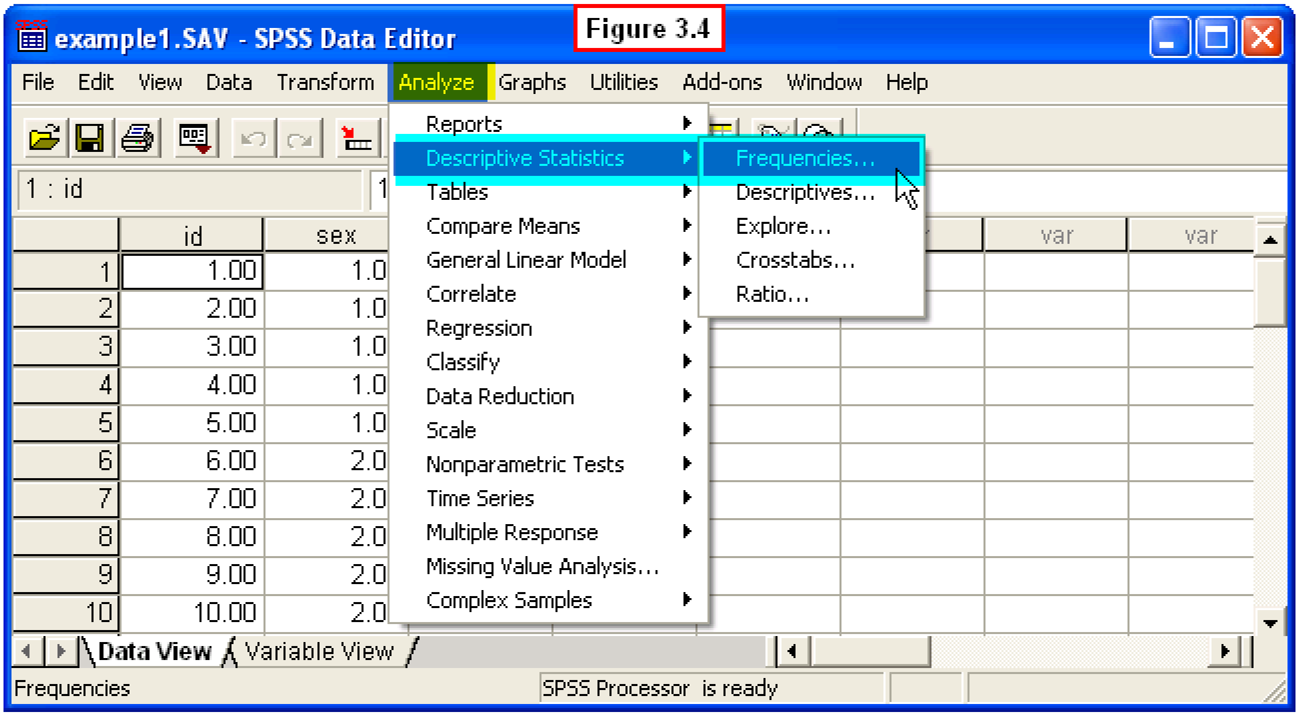


Recall that this file contains data for ten students in a statistics course, including an **ID** number, a code for **sex** (1: male; 2: female), and the number of previous psychology **courses** taken.

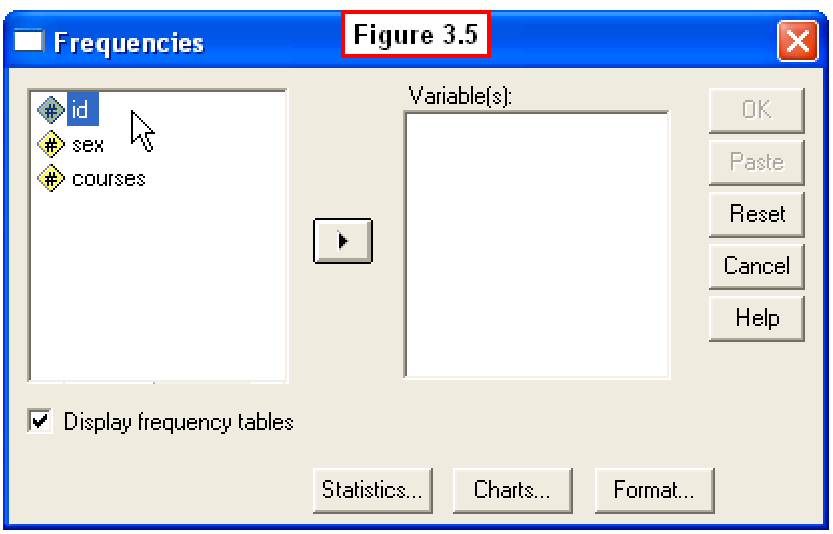
Show Me Video!

3.1b Running the procedure: the Frequencies dialog box

To run the frequencies procedure select **Analyze, Descriptive Statistics, Frequencies…** from the drop-down menu (Figure 3.4).



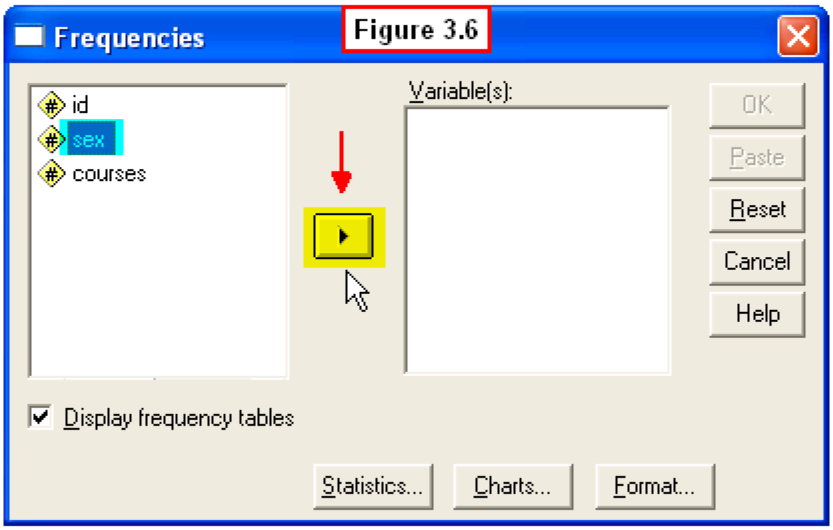
A **Frequencies dialog box** will appear listing our variables in the pane on the left (Figure 3.5).



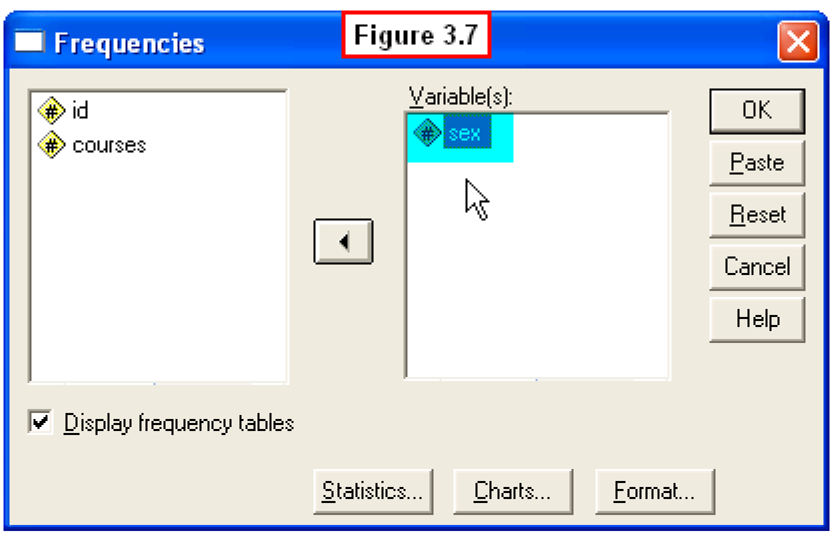
In this box, we select which variables we want to analyze. This will be done by highlighting the desired variables on the left, then moving them into the **Variable(s):** pane on the right.

The **id** variable is initially highlighted. However, we would generally not be interested in generating frequency tables for this variable, so move your cursor down and click on the **sex** variable in the left panel.

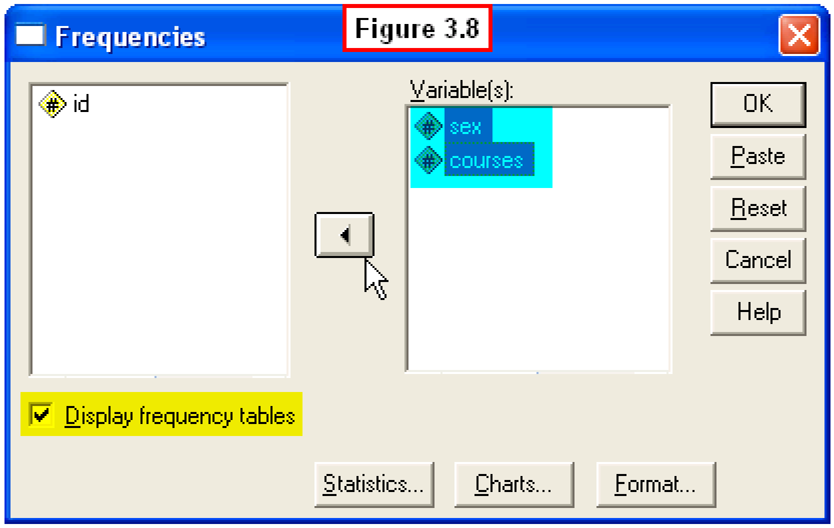
Now that we have selected this variable, we need to move it into the **Variable(s)** choice pane on the right. To move this variable, click the arrow key in the middle of the dialog box (Figure 3.6).



When you have done this, the variable will now appear in the right pane (Figure 3.7).



Now click on the **courses** variable in the left pane and move it into the **Variable(s)** selection pane. When you are done, your Frequencies dialogue window will look like the one in Figure 3.8.

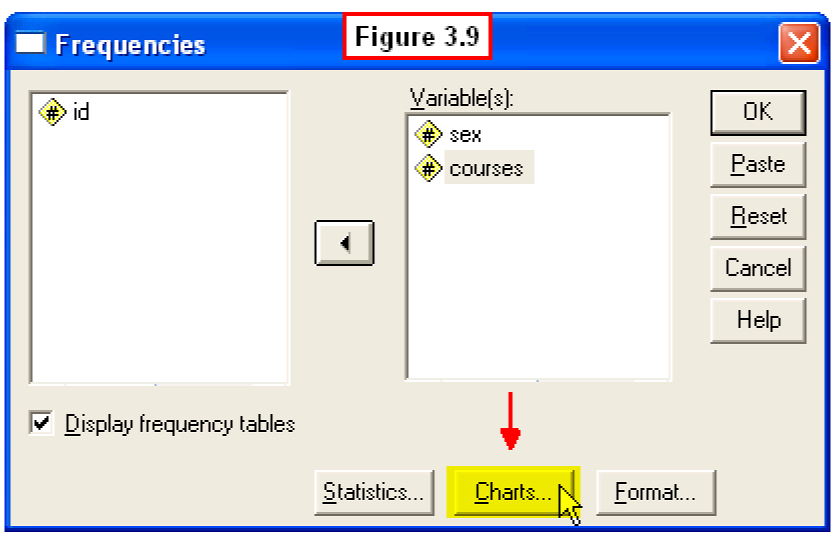


Note that by default, SPSS has checked the box labeled **Display frequency tables** in the lower left. This will result in the generation of frequency tables of our selected variables when we are ready to run this procedure.

Show Me Video!

3.1 c The Frequencies: Charts dialog box

So far all we have done is specify which variables for which frequency tables will be generated. However, the interpretation of frequency tables is often facilitated by generating charts of the frequencies, so we will turn to that process before running the procedure. To request charts, click the **Charts** button at the bottom of the **Frequencies** dialog box (Figure 3.9).



This causes a new **Frequencies: Charts** dialog box to appear (Figure 3.10).

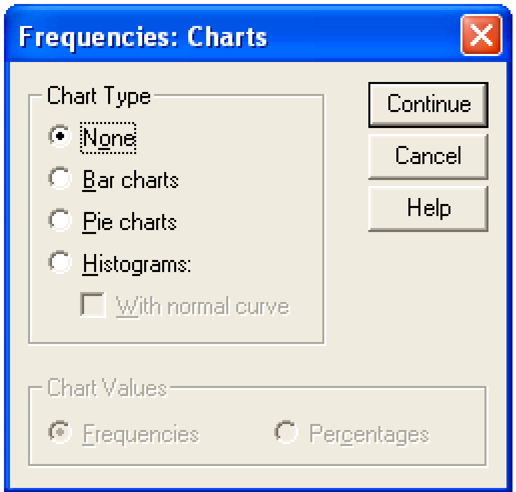


Figure 3.10

This box allows us to select which type of chart to create. For this example, click the **Bar charts** radio button. We could also choose percentages instead of frequencies as the **Chart values** to be displayed, but we will leave the frequencies check box selected. Now click the **Continue** button in the upper right corner of this dialogue window (Figure 3.11).

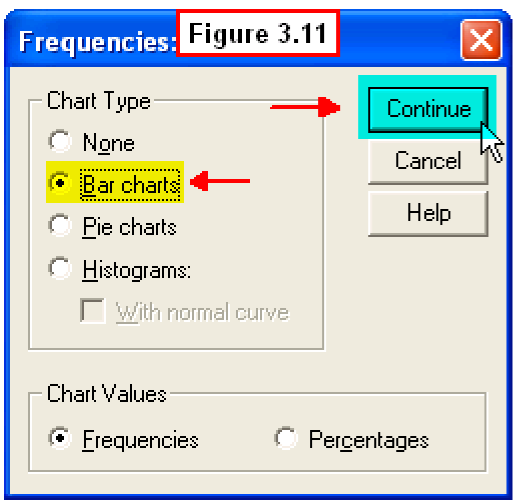
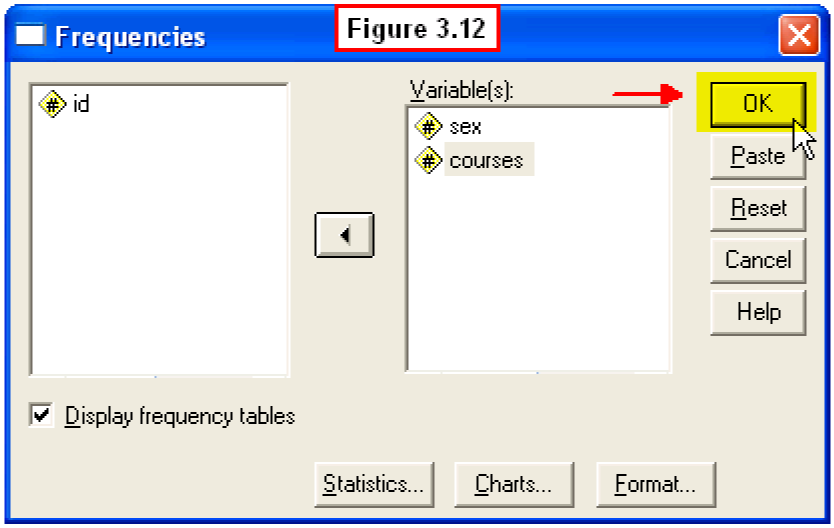


Figure 3.11

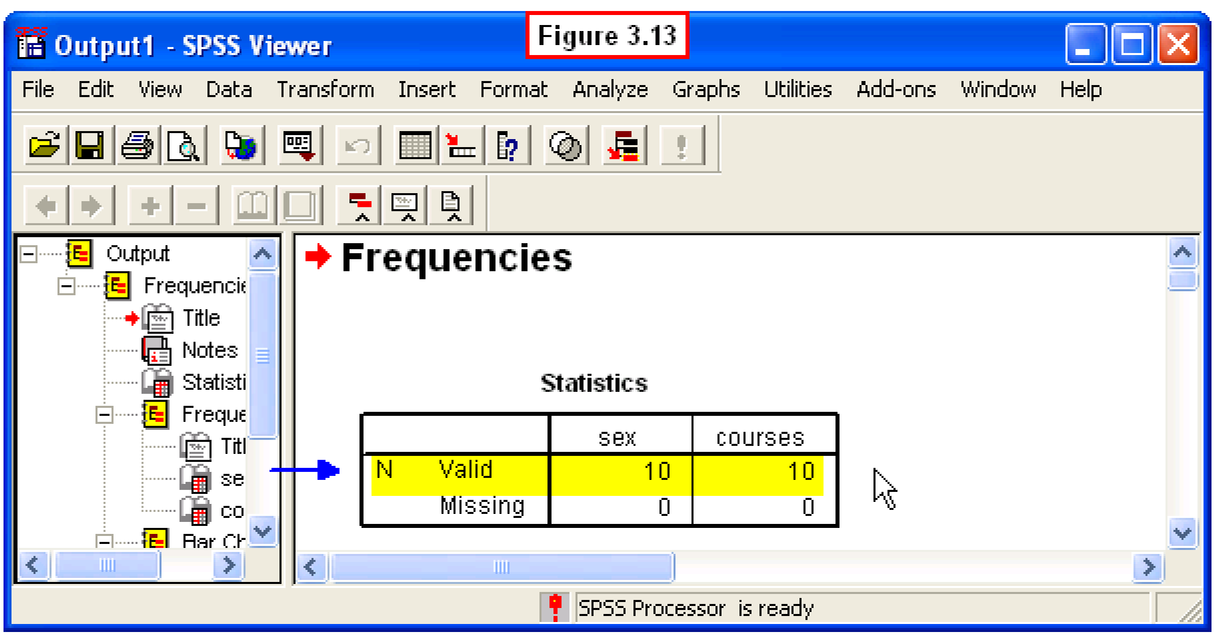
The **Frequencies: Charts** box will close, returning you to the **Frequencies** dialogue box. Now we are ready to run this procedure. To do this, click the **OK** button in the upper right corner of this box (Figure 3.12).



[Show Me Video!](#)

3.2 Interpreting the Frequencies Procedure Output

The results of this procedure will now appear in an SPSS Output Viewer window (Figure 3.13).



The first information in the output is the **Statistics** table. This table displays how many valid cases (**N**) were processed and how many cases had missing values for each of our variables. Since we have no missing values, the number of valid cases is the full 10 students for both variables.

Note: Rather than showing the entire viewer window as we discuss various parts of output files, hereafter we will display different parts of the output as separate figures. The next table is the simple frequency distribution for the variable, **sex** (Figure 3.14). To find this table, just scroll down your output viewer window to find this table, or click that item in the tree menu in the left pane.

Figure 3.14

sex

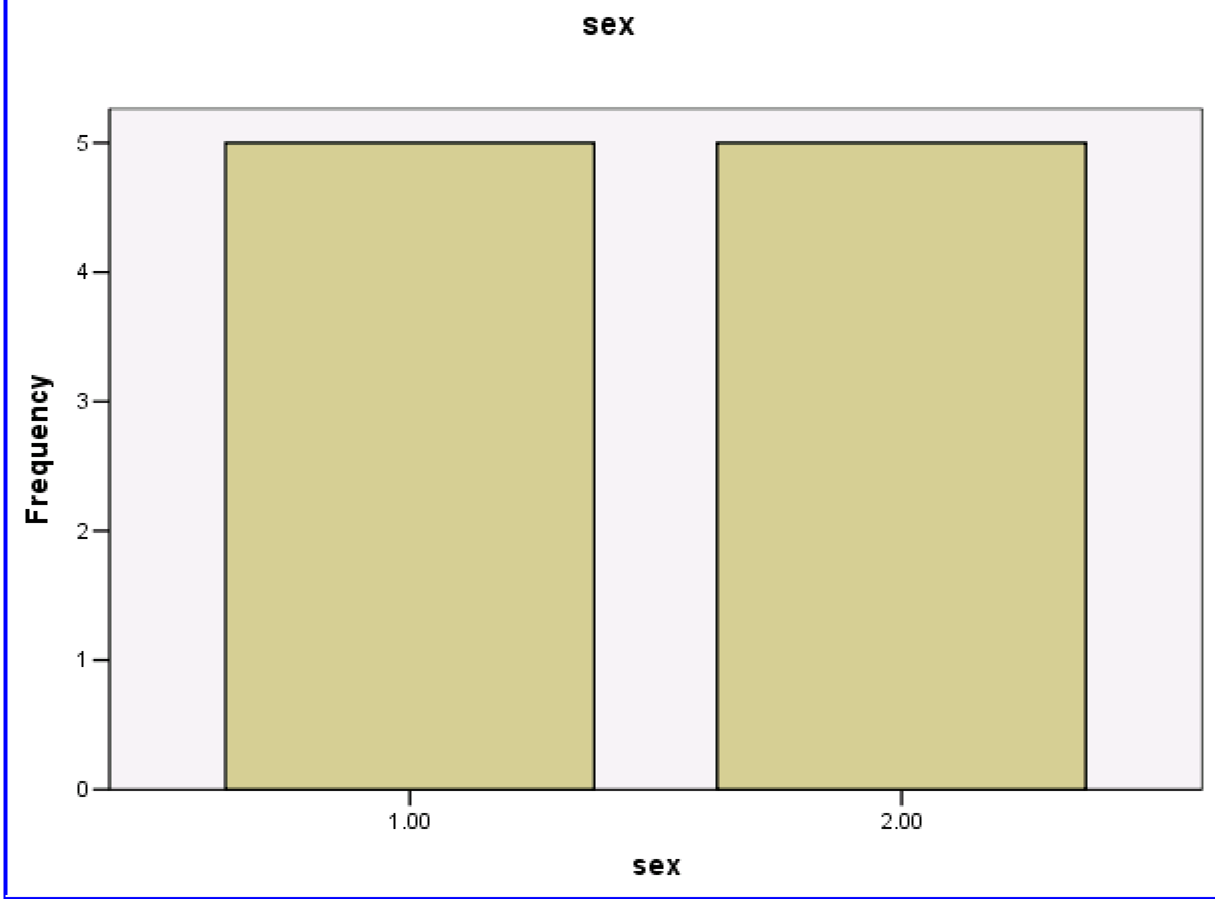
1=male; 2=female

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	5	50.0	50.0	50.0
	2.00	5	50.0	50.0	100.0
	Total	10	100.0	100.0	

The first column lists the labels we assigned to the two levels of this variable (1: male; 2: female). The **Frequency** column displays the frequency of each score (in this case, category). This shows that of the ten students, five were women and five were men. These frequencies are converted to percentages in the **Percent** column (50% men, 50% women). Note the **Valid Percent** column shows the same values. These would be different if we had missing data; i.e., this column adjusts the percentages based on missing values.

Scroll down your output and you will see the bar chart we generated for this variable (Figure 3.15).

Figure 3.15



This lists the numerical values of the variable (1: male; 2: female) on the horizontal axis and the frequencies (how many instances of each sex) on the vertical axis. This visual depiction of the results clearly shows an equal number of men and women, with the two bars of the chart being the same height.

Scroll back up to see the next table presenting the distribution of scores on the **courses** variable (Figure 3.16).

Figure 3.16

courses

# courses		Frequency (# students)	Percent	Valid Percent	Cumulative Percent
Valid	1.00	2	20.0	20.0	20.0
	2.00	3	30.0	30.0	50.0
	3.00	4	40.0	40.0	90.0
	4.00	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

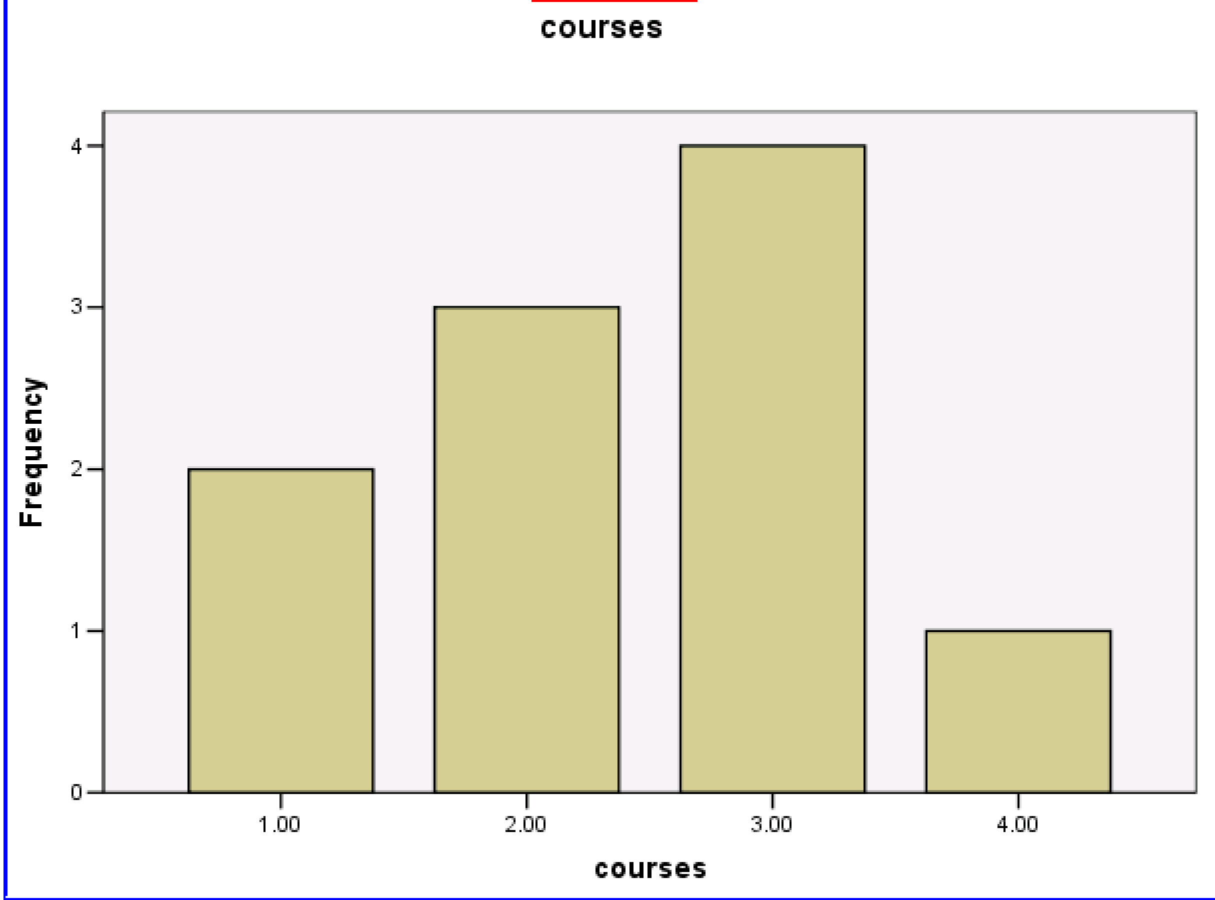
In this table, the scores in the first column are the number of psychology courses taken by the students. Thus, students had taken between 1 and 4 courses prior to statistics.

The **Frequency** column shows how many students had 1, 2, 3 or 4 courses. Thus, two students had taken 1 course, three had taken 2 courses, four had taken 3 courses, and one had taken 4 courses.

Again, these frequencies are converted to percentages in the **Percent** column. From this inspection, we can conclude the most students had taken either 2 psychology courses (30%) or 3 (40%) courses.

Scroll down and you will see the bar chart of this frequency distribution (Figure 3.17).

Figure 3.17



The bar chart again provides a visual depiction of this distribution. A glance at this graph clearly shows that the most frequent number of courses taken by students prior to taking statistics was 3, followed by 2 courses.

3.3 A Look Back and a Look Ahead

Chapters 2 and 3 were meant to get you a quick start with SPSS, to illustrate the basics of SPSS, creating a data file, and doing a simple analysis of the data in that file. The **Frequencies** procedure illustrated is commonly used, especially with survey research. It is an effective number cruncher for summarizing data (and is especially important with large data files). For example, the frequency distribution of the number of previous psychology courses taken by students can be useful to the instructor in understanding the variability in the psychology background of his/her students.

There are other analyses that can be generated using the **Frequencies** procedure (e.g., we could generate descriptive statistics, such as the mean or mode). We could also generate separate frequency tables (e.g., to compare the number of course taken by male vs. female students). We will return to the **Frequencies** procedure in a later chapter.

Now that you have a basic familiarity with SPSS, in the next chapter we will introduce the **EZDATA** file that we will be using for the remainder of the text. This file is much larger and more complex than the simple file we have been using so far.

Thus, we will devote an entire chapter to explaining the variables in this file and how to import it into SPSS. This is an important chapter to read and study, since it will be used in the remaining chapters. Spending some time now to understand the logic of the **EZDATA** research project and the variables included in the data file will save you time later in the interpretation of the data analyses that will be performed on this file.

3.4 Chapter Review Video

Review Me!

3.5 Try It! Exercises

1. Running the Frequencies Procedure in the Data Editor

Open the **example1.sav** file in the Data Editor. Use the **Frequencies** procedure described in Section 3.1 to generate the output file shown in Section 3.2.

- **Print** this file to submit to your instructor by selecting **File, Print** from the Output Viewer menu.

Chapter 4
Variables in the EZDATA File: The Sex Roles, Work Motives & Leadership Project

4.1 Before we create the data file...

By now you should have a basic understanding of SPSS, and know how to create a data file and how to generate simple frequency tables. We are now ready to describe a set of data that will provide the basis for statistical analyses using SPSS for the remainder of this book. We will devote this entire chapter to describing this data file obtained from a hypothetical study of sex roles and leadership effectiveness.

In creating a data file, a researcher must 1) decide which variables to include, 2) have measurements of these variables (i.e., scores) from participants, 3) organize these scores for input to the file, and 4) enter the scores into a data file to be saved for later use. Were we to carry out an actual research project, these steps could be both involved and time consuming (especially steps 1 and 2, which require reviewing the literature, designing the research project, and carrying it out).

We plan to save time by describing a hypothetical research project for which we have already generated data. We created the results to reflect those that might be expected if the study had actually been conducted. Thus, in this chapter we will present you with a set of variables and how they were measured to be entered as data from this hypothetical project. In Chapter 5 we will provide the actual scores for this data file and show you how to enter this data file into SPSS.

You will then use this data file to conduct a variety of statistical analyses using SPSS. These procedures will be introduced in the remaining chapters of this book, and you will be asked to apply them to this data file to analyze and interpret of the results of the hypothetical project.

4.2 A Hypothetical Project on Leadership Effectiveness

In the following sections, we first discuss the rationale behind this project and explain the variables to be included. We will hint at how the results of this study might turn out, but we stop short of a full explanation. Thus, even after reading the variable descriptions and creating the data file, you may be uncertain about the way in which the data relate to the purpose of this project.

But that's alright, as many researchers feel the same way after actually having completed a study! That's what statistical analyses are for – helping the researcher understand the data – and part of the excitement of research comes from the gradual unfolding of the results of the study through data analyses. You will find that SPSS is a very powerful tool which can assist you in this discovery process.

The data from this hypothetical project have been fabricated to reflect some of the actual research in this area. The results have been created to be meaningful, and many may confirm your intuitions. So while you may find the study confusing at first, you should gradually develop a good understanding of the project as you progress through the exercises in this book.

Suppose that you are a social scientist who has been hired by a large corporation (EZ Manufacturing, Inc.) to conduct a study of leadership effectiveness. EZ is a high tech electronics manufacturing company which employs several thousand men and women on the assembly line, and is directed by several hundred people, mostly men, at various levels of management.

The upper-level management of this organization is interested in obtaining information relevant to their planned affirmative action program of promoting qualified women to management positions. While they are enthusiastic about this program, there is some apprehension due to the fact that these positions have traditionally been occupied almost exclusively by men.

4.3 Overview of Major Variables

EZ Manufacturing execs have asked you to draw upon research in this area to conduct your study in their organization, hoping that your research will help them determine what types of individuals (both men and women) are most likely to be effective leaders. You identify two areas of research that have been related to leadership effectiveness:

- **Gender and Sex Role Stereotypes**
- **Leadership Style and Work Motives**

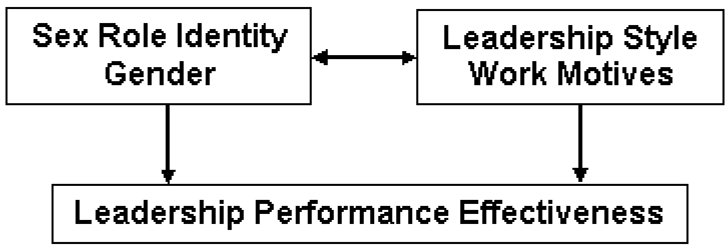
In a general sense we can think of these as ***predictor variables***. That is, we may want to see if these variables predict some criterion, or ***outcome variable***. The outcome variable we want to predict is **Leader Performance Effectiveness**., so we are interested in discovering how these variables might be related to effective leader behavior.

The first area, **Gender** and **Sex Roles**, *concerns the extent to which men and women internalize societal stereotypes about masculinity and femininity in their self-concepts*. This research is relevant to your planned project in leadership positions have traditionally been traditionally occupied by men and have been associated with masculine personality characteristics.

The second topic, **Leadership Style** *concerns the extent to which a person exhibits task skills and/or social skills in leadership situations.* The third area, **Work Motives**, *concerns the extent to which a person strives to fill needs for affiliation, achievement and dominance in the work setting*. Both leadership style and motivation are directly relevant to how well a leader performs.

Thus, you determine that it might be useful to investigate all of these variables and their interrelationships as predictors of leadership performance at EZ Manufacturing. Figure 4.1 graphically depicts the interrelationships among these variables.

Figure 4.1



The two-way arrow between the sets of predictor variables indicates that we will be examining interrelationships among them. The one-way arrows indicate that we will be investigating how leader performance varies as a function of the two sets of predictor variables. In the next section we discuss the operational definitions of these variables, and explain how they were measured.

4.4 The First Set of Predictor Variables: Sex Role Identity and Gender

As is often the case in research projects, we will obtain more than a single measurement related to sex role identity (how masculine or feminine a person perceives him/herself). In fact, we shall rely on 10 items on a questionnaire to assess this concept.

Further, one's sex role identity is often related to one's gender - you might suppose, for example, that men are more likely to be masculine types, and women feminine types. As you will soon see, however, one's gender and sex role identity, though related, are not one and the same. Thus, we will need to measure gender as well as sex role identity.

4.4a Sex Role Stereotypes & Self Identity

Sandra Bem (1972) and others (e.g., Eagly, 1990) have conducted numerous investigations indicating that male and female traits, roles and behaviors in our culture are typically perceived in rather different and stereotypic ways. For example, men are generally thought to be assertive, independent, aggressive, decisive and unemotional, while women are seen as sympathetic, compassionate, understanding, nurturing, and emotional.

Of course, these are stereotypes of men and women, and it is possible for a woman to be assertive and a man to be compassionate. However, studies have shown that most men and women perceive themselves in stereotypic ways. That is, a ***Sex-typed*** man would have a self-concept consisting of primarily masculine traits, and a sex-typed woman would perceive primarily feminine traits in herself.

Despite the fact that most people think of themselves (and others) in stereotyped ways, Bem identified another category of men and women that she termed ***Androgynous*** in their sex role identity. Androgynous men and women are individuals who perceive themselves to have ***both*** masculine and feminine characteristics. Thus, for example, an androgynous man would see himself as *both* assertive and sympathetic, and an androgynous woman would see herself as *both* independent and compassionate.

4.4b Measuring Sex Role Identity & Gender

A variety of methods have been used to assess the degree to which an individual is sex-typed vs. androgynous in her/his self- concept. Bem (1972) devised the Bem Sex Role Inventory (BSRI), consisting of twenty masculine and twenty feminine personality traits. Respondents indicate the extent to which these traits describe themselves. Rather than use the entire BSRI instrument, we will include measures of employees' self-ratings for only five of the masculine and feminine traits from the BSRI (see Table 4.1).

Table 4.1	
Masculine Traits	Assertive; Decisive; Independent; Self-reliant; Aggressive
Feminine Traits	Compassionate; Nurturing; Sympathetic; Understanding; Emotional

Since it is best to use short, simple variable names in SPSS, we will simply call these trait self-ratings **masc1** through **masc5** and **fem1** through **fem5**.Thus, the first ten variables in our data file will be employees' self-ratings on these five masculine and five feminine traits. These trait scores can range between 1 and 7, where

- **1** = **Not at all Descriptive of Me** (meaning low femininity or low masculinity)
- **7** = **Completely Descriptive of Me** (meaning high femininity or high masculinity)

As we will see later, **Sex Role Identity** will be assessed by a composite of scores on these ten traits. Specifically, a person's sex role identity will be determined by the extent to which employees score low or high on the masculinity and feminity scales.

It may have occurred to you that the employees' **gender** is an important variable to consider in recording the employee feminine/masculine trait scores. For example, we cannot assume that a high femininity score necessarily means that the individual is a woman.

Indeed, as we have seen, it is possible for both men and women to score high on the femininity. By the same token, both men and women can score high on masculinity (recall that androgynous individuals score high on both masculinity and femininity).

Thus, in addition to recording a given employee's femininity/masculinity scores, it is important to record that person's gender. This is accomplished easily enough by asking each employee to circle a 1 at the top of the assessment questionnaire if he is a man, and a 2 if she is a woman. Thus, assume that the 11th variable you measure is **Gender**, where:

- **1** = **Male employees**
- **2** = **Female employees**

You might begin thinking ahead to how the variables of sex role identity and gender are related to leadership style, work motives and performance effectiveness. Of course, the remainder of this text will explore some of these relationships.

4.5 The Second Set of Predictor Variables: Leadership Style & Work Motives

Below we discuss the research relating to the second predictor, **Leadership Style**, and describe the measures used to assess leadership style. Then we will turn our attention to **Work Motives**.

4.5a Defining & Measuring Leadership Style

Research on leadership effectiveness has indicated that people differ in their preferred **Leadership Style**. The two main categories of leadership style are **Relations** orientation and **Task** orientation (Feidler, Chemers, & Mahar, 1976). ***Relations-oriented*** leaders gain satisfaction from interpersonal relationships, while ***task-oriented*** leaders gain satisfaction from task accomplishment.

Relations-oriented leaders attempt to maintain high productivity by promoting good interpersonal relationships among subordinates, whereas task-oriented leaders attempt to maintain productivity by arranging working conditions such that the human element interferes to a minimal degree. Thus, relations-oriented leaders have strong **social skills**, while task-oriented leaders have strong **task skills**. Examples of the types of behaviors representing these skills are shown in Table 4.2.

Table 4.2	
Social Skills	Effective listening; resolving conflicts; focus on worker needs
Task Skills	Effective decision making; meeting deadlines; focus on quality

Research shows that each of these leadership styles is likely to be effective in some situations, but ineffective in others. However, just as we saw with masculinity/femininity, even though many managers are either primarily relations or task oriented in their leadership styles, Blake & Mouton (1980) suggested that it is possible for a person to exhibit high levels of *both* social skills and task skills in her/his leadership style.

Blake and Mouton (1980) developed a questionnaire that yields separate scores for a person's social skills and task skills. To simplify, we will assume that EZ employees completed this instrument, and each employee received a score from 1 to 9 indicating his/her degree of social and task skills. Thus, the next two variables in our data file will be **Social Skills** and **Task Skills** (which we will name **soc** and **task** in the file).

The range of scores on the **soc** variable is as follows:

- **1** = **Low Social Skills**
- **9** = **High Social Skills**.

The same scoring system will be used to measure **task**:

- **1** = **Low Task Skills**
- **9** = **High Task Skills**

If you have been thinking about the possible connections among the variables described above, good for you - you are showing the curiosity that makes for a good researcher! For example, you may have considered the possibility that masculine sex-typed employees might score high on task skills and low on social skills, while feminine sex-typed individuals might score low on task skills and high on social skills.

Further, you may have anticipated that the male employees at EZ Manufacturing are likely to perceive themselves as task-oriented, while female employees are likely to see themselves as relations-oriented. These intuitive predictions follow from cultural stereotypes regarding the personality and behavior of men and women.

However, you might also think about how androgynous men and women score on the task and relations dimensions of leadership style. Only the data from your study can suggest the accuracy of your predictions, and we will test such hypotheses in subsequent chapters of this book.

4.5 b Defining & Measuring Work Motives

Research in organizations indicates that productivity (and leadership potential) can also be understood in terms of the relative importance of various needs people strive to satisfy on the job. Examples of work motives include ***achievement*** needs (the desire to accomplish goals and be recognized for accomplishments), ***affiliation*** needs (the desire for rewarding interactions with co-workers) ***dominance*** needs (the desire to exert power and influence on others).

Steers and Braunstein (1976) developed a measure of these work motives. Respondents indicate how frequently each of several behaviors relevant to satisfying the above needs applies to their behavior on the job. See Table 4.3 for examples of work behaviors reflecting these needs.

Table 4.3	
Achievement	I try very hard to improve my past performance at work.
Affiliation	I find myself talking to others about nonbusiness-related matters.
Dominance	I strive to be in command when I am working in a group.

Participants rate themselves on each behavior using a scale from 1 (never true) to 7 (always true). Assume that you have obtained self-ratings from EZ employees on five behaviors relevant to each of the above three areas of work motivation. We will simply name these variables **ach1** through **ach5** (achievement needs), **aff1** through **aff5** (affiliation needs) and **dom1** through **dom5** (dominance needs).

Scores on each of these 15 work behaviors are as follows:

- **1** = **Never True of Me** (meaning a low level of that need)
- **7** = **Always True of Me** (meaning a high level of that need)

You might take a moment to think about the possible relationships between these work motivation dimensions and the previous variables. For example, it might be that task-oriented leaders tend to score high on achievement needs, while relations-oriented leaders score high on affiliation needs.

Cultural stereotypes exist regarding differences between men and women in the above needs, so you might think about what the results will show regarding gender and sex-role identity differences in needs. Again, these issues will be addressed in later chapters via the various statistical procedures you will learn to conduct using SPSS.

4.6 The Major Outcome Variable: Performance Effectiveness

In most instances, we can think of the previous variables discussed as the predictor or independent variables in our study. And, in general, we are interested in how these predictor variables impact or relate to a particular outcome variable, leadership effectiveness. However, as you will see, in some instances it may be beneficial to investigate relationships among various predictor variables themselves.

But for now, our task is to consider the major outcome or dependent variable in our project. We are interested in observing differences in this variable as a function of the many predictor variables mentioned above, so that we can implement the policy of promoting individuals within the company who are likely to become effective leaders.

We would need to begin by obtaining measures of **Performance Effectiveness** for each employee in leadership situations. This could be a rather complicated process, but to simplify things, let's assume that you have asked each employee's supervisor to examine this person's performance within a variety of leadership situations during a six month period. The supervisors are asked to give an overall rating (from 1 to 9) of the employee's performance during these six months, yielding your outcome variable, which we will name **perform** in our data file.

Scores on the leadership **perform** variable range from:

- **1** = **Not at All Effective**
- **9** = **Extremely Effective**

Many of the SPSS analyses which we will conduct throughout this text will use this variable as the outcome variable to be related to differences along the predictor variables described above. However, we will sometimes treat the other variables as dependent variables themselves when examining their interrelationships.

4.7 Repeated Measures Variables: Social Skills, Task Skills & Performance

Assume that in addition to examining interrelationships among the variables measured to identify employees with leadership potential, EZ execs also have asked you to develop a management training program to improve the leadership skills and performance of the employees in your study. Here also we will simplify and follow Blake and Mouton's (1980) suggestion that it is possible for all people to develop both social and task skills related to effective leadership. So the focus on your week-long training program is on improving employees' skills in both of these areas of leader behavior.

Since it is always important to assess the effectiveness of programs such as this implemented in an organization, you decide to rely on variables you have already measured to evaluate this management training program. Specifically, since both task and social skills are important in leadership, if your program is successful, we should see an increase in both sets of skills after participation in the program. Further, we would also expect to see increases in actual leader performance after the program.

Since you already obtained scores from employees on **soc**, **task**, and **perform** at the beginning of your study, you can assess your program's effectiveness by re-measuring employees on these variables after completing the workshop. The same scales and scoring procedures for these three variables will be used at the second measurement, but we will need to give these new scores different variable names. We will name them **soc2**, **task2**, and **perform2** in the data file.

Another typical procedure in program evaluation is to obtain immediate and delayed assessments to see if any improvements are long-lasting. Thus, in addition to re-measuring **soc**, **task**, and **perform** immediately after participation, assume that you re-administer these scales three months later. Here also, we will need to give these new scores different variable names, which will be **soc3**, **task3**, and **perform3**.

Researchers refer to these as **repeated measures variables**, because we are obtaining more than one measurement of the same variable at three different times (e.g., **perform1, perform2** and **perform3**). These types of variables must be treated differently than single-measurement variables when conducting statistical analyses. We will explain this in greater detail in a later chapter.

4.8 Summary of Variables in the Data File

Table 4.5 summarizes the measured variables and score ranges that will comprise your data file.

Table 4.5	
Variable	**Scores**
GENDER	1 = Men; 2 = Women
MASC1-MASC5	1 = Low Masculinity; 7 = High Masculinity
FEM1-FEM5	1 = Low Femininity; 7 = High Femininity
SOC1-SOC3	1 = Low Social Skills; 9 = High Social Skills
TASK1-TASK3	1 = Low Task Skills; 9 = High Task Skills
ACH1-ACH5	1 = Low Achievement Needs; 7 = High Achievement Needs
AFF1-AFF5	1 = Low Affiliation Needs; 7 = High Affiliation Needs
DOM1-DOM5	1 = Low Dominance Needs; 7 = High Dominance

	Needs
PERFORM1-PERFORM3	1 = Not at All Effective; 9 = Extremely Effective

The data file itself will consist of the scores each employee received on each variable. **Note:** familiarize yourself with these variables so you will understand the file to be described in Chapter 5. More importantly, beginning with Chapter 6 you will be asked to write interpretations of the data analyses you perform. If you do not understand these variables, you will not be able to write interpretations of your analyses!

For example, you will be asked to describe the distribution of scores on the **soc** variable at EZ. If you do not know that a score of 1 on the **soc** variable means low social skills in leadership style and a 9 indicates high social skills, you will be unable to write a meaningful interpretation. So we recommend that you do yourself a favor and spend some time now studying the variables in Table 4.5.

Chapter 5
Housekeeping:
Transforming Variables and Adding Labels

5.1 The EZDATA File

In this chapter we provide the EZDATA file for you to download. We also describe it's structure and organization, and we explain modifications of the original file that have been made to facilitate it's use in analyzing the results of the leadership project.

5.1a Downloading the EZDATA file

Click the **EZDATA button** below to download the EZDATA file we have been describing and will be using for the remainder of this text. Downloading and opening this file will facilitate your understanding as we describe its structure and organization. When you click the button below you will be prompted to either open the file or save it.

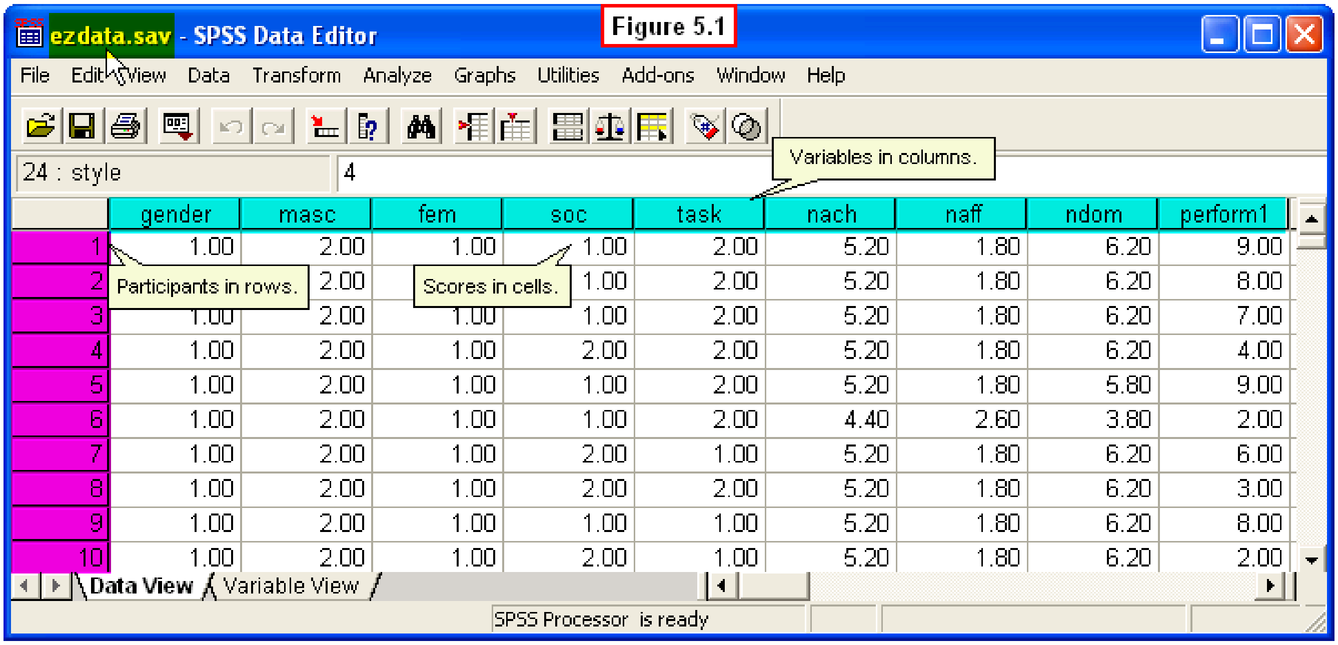
Select the save option, then navigate to the ezdata folder you created for storing files. Name the file **ezdata.sav** for easy recall.

Download EZDATA

Show Me Video!

5.1b Structure of the EZDATA file

After you have downloaded the file, open it in the SPSS Data Editor. When you have opened the file in SPSS, it should look like the one shown in Figure 5.1.



As discussed in Chapter 2.3b, the data in the file are organized with all of our **variables** in the **columns**, and individual **employee participants** (cases) listed in **rows**. The **scores** each employee received on each of the variables are entered in the **cells** along each row.

For example, employee number 1 was assigned a score of 1 for the **gender** variable (to indicate he is a man); he received scores of 1 on the **masc** variable and 2 on the **fem** variable. Scroll to the right and you will see his scores on the other variables in the data file.

As you continue scrolling to the right, you may notice that the listing of the variables in the columns looks different from the listing explained in Chapter 4. You may have also noted that there are some new variables in the file that weren't discussed in Chapter 4. The reason for this is that we have already done some of the housekeeping referred to in the title of the present chapter.

That is, we have completed transformations on the original variables to create the new variables in the file, and have added variable labels for the major variables, as well as value labels for some of them. We did this so that we could provide the file in its complete form for download. This way we can be sure you will be working with exactly the same **ezdata.sav** file that we will be using in the remainder of this book. Also, this should reduce any anxiety you might have about losing or somehow ruining your ezdata file - you can come back to this page anytime and re-download the file.

In this chapter we will explain the procedures for making these changes to the file. Some of them are tedious and error-prone, which is why we have already done them. In the remaining chapters of this book, we *do* suggest that you actually follow along examples and perform data analysis procedures on the **ezdata.sav** file that we describe.

But in the current chapter, instead of asking you to perform the file modification procedures yourself on the **ezdata.sav** file, we will suggest that you practice these procedures on the simpler **example1.sav** file you created in Chapter 2. That way you will still learn the procedures without having to perform them all on the **ezdata.sav** file.

5.1 c So what are these new variables?

If you look back at Table 4.5 and also scroll through the **ezdata.sav** file in the SPSS Data Editor, you will see that the original variables shown in Table 4.5 do exist in the new ezdata file, they are just in different columns than you might have expected. Gender does appear in the first column, but the **masc1-masc5** variables (the self-ratings on the 5 masculine personality traits) do not begin in the second column. Instead, scroll right and you will see them in columns 18-22. The **fem1-fem5** scores appear in columns 23-27.

The variables named **masc** and **fem** that you see in columns 2 and 3 are actually new variables that have been created by transformations of the original **masc1-masc5** and **fem1-fem5** variables. Click the **Variable View** tab at the bottom of the Data Editor for a better view of all of the variables. Recall that this view lists variables in rows instead of columns (Figure 5.2).

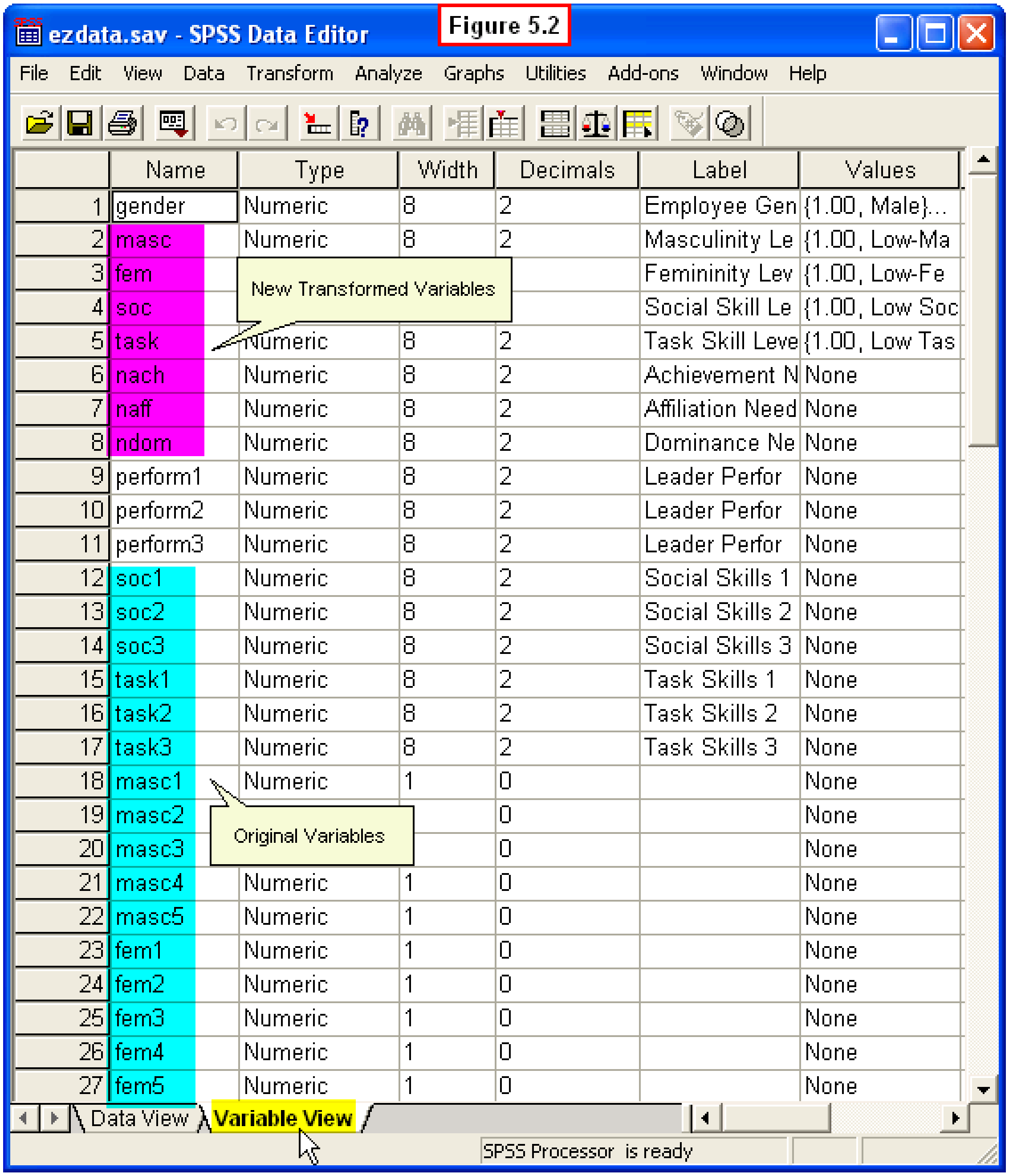


Figure 5.2

To create the new variables, **masc** and **fem**, we used the SPSS **Transform** procedure to add up the scores on the **masc1-masc5** and **fem1-fem5** variables to create two variables named **masctot** and **femtot**. If you scroll all the way down, you will see these masctot and femtot variables (SPSS tacks newly-created variables at the end of the original data file).

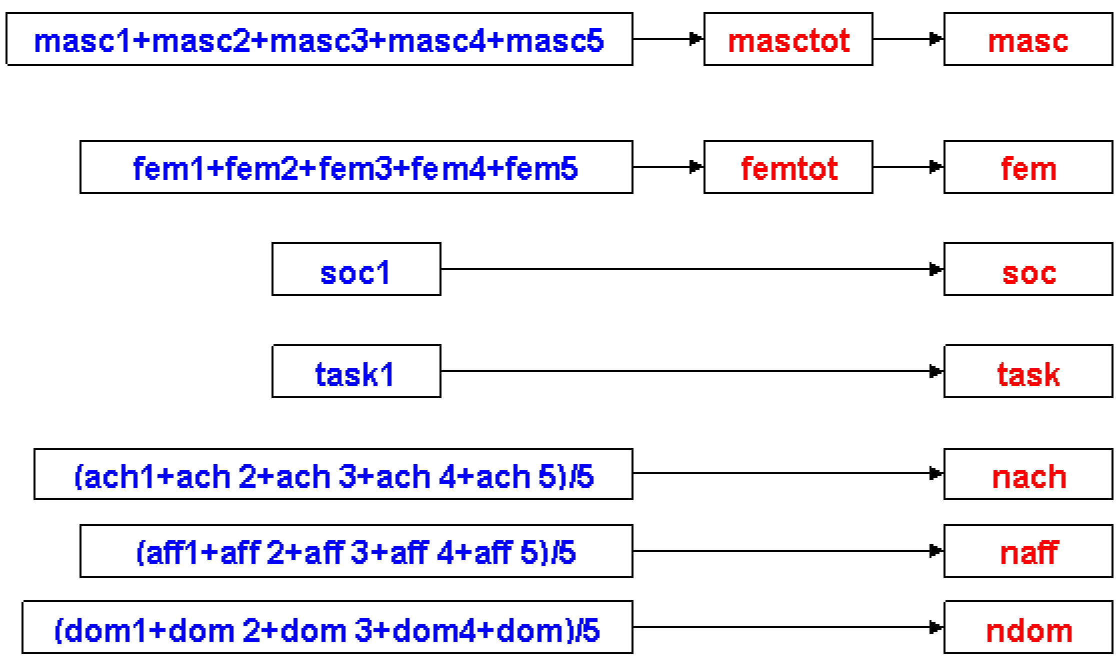
Next, we used the SPSS Transform procedure again to convert the continuous masctot and femtot variables into categorical variables, which we named **masc** and **fem**. Last, we moved these two new variables to rows 2 and 3 for easy access later. If these variable transformations sound confusing, don't worry - we will explain all of this in the next section!

You will also see the original **soc1-soc3** and **task1-task3** variables (employee's scores on social and task leadership skills) in rows 12-14 and 15-17, respectively. We performed similar transformations on the **soc1** and **task1** variables to create new categorical variables that we named **soc** and **task**, and then placed them in rows 4 and 5 for easy access.

Next, you will see the original **ach1-ach5**, **aff1-aff5**, and **dom1-dom5** variables (employee work motive scores) in rows 28-42. We used the the Transform procedure on these variables to add them up and take their average to create the new variables, **nach**, **naff**, and **ndom** (short for achievement needs, affiliation needs and dominance needs), and we placed these three new variables in rows 6-8 for quick access.

Figure 5.3 graphically depicts these variable transformations. Note that the original variables (in blue) still exist in the data file, and their transformed versions have been given new names (in red), and now exist as separate new variables in the file.

Figure 5.3



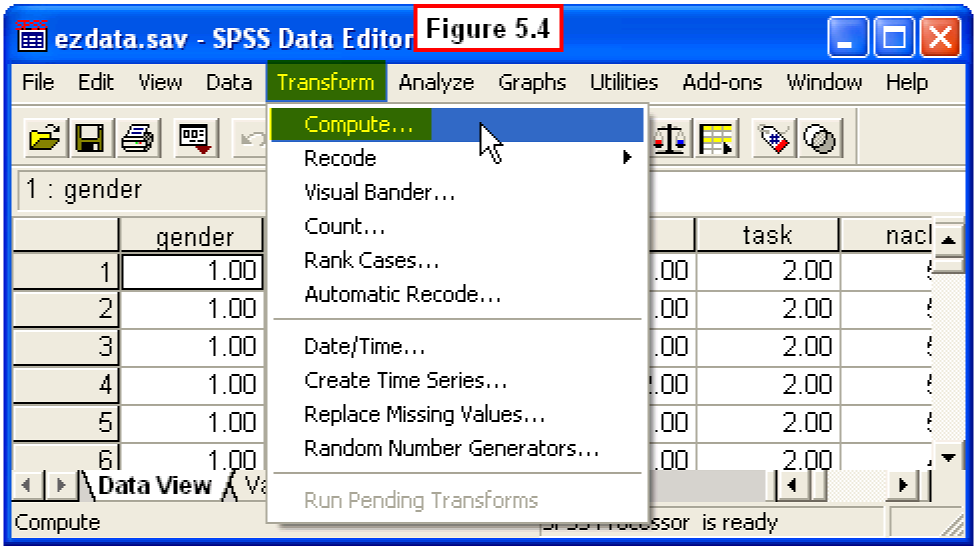
You will find the original **perform1-perform3** variables in rows 9-11. We have not done any transformations on these variables. Last, if you scroll all the way down, you will see yet two more new variables, **sextype** and **leaderstyle**. We also used the Transform procedure to create these new categorical variables, but we will not discuss how we did this in the present chapter. We explain this procedure in a later chapter since it is a bit more complex than the others we describe in this chapter.

5.2 Creating New Variables using the Transform Procedure

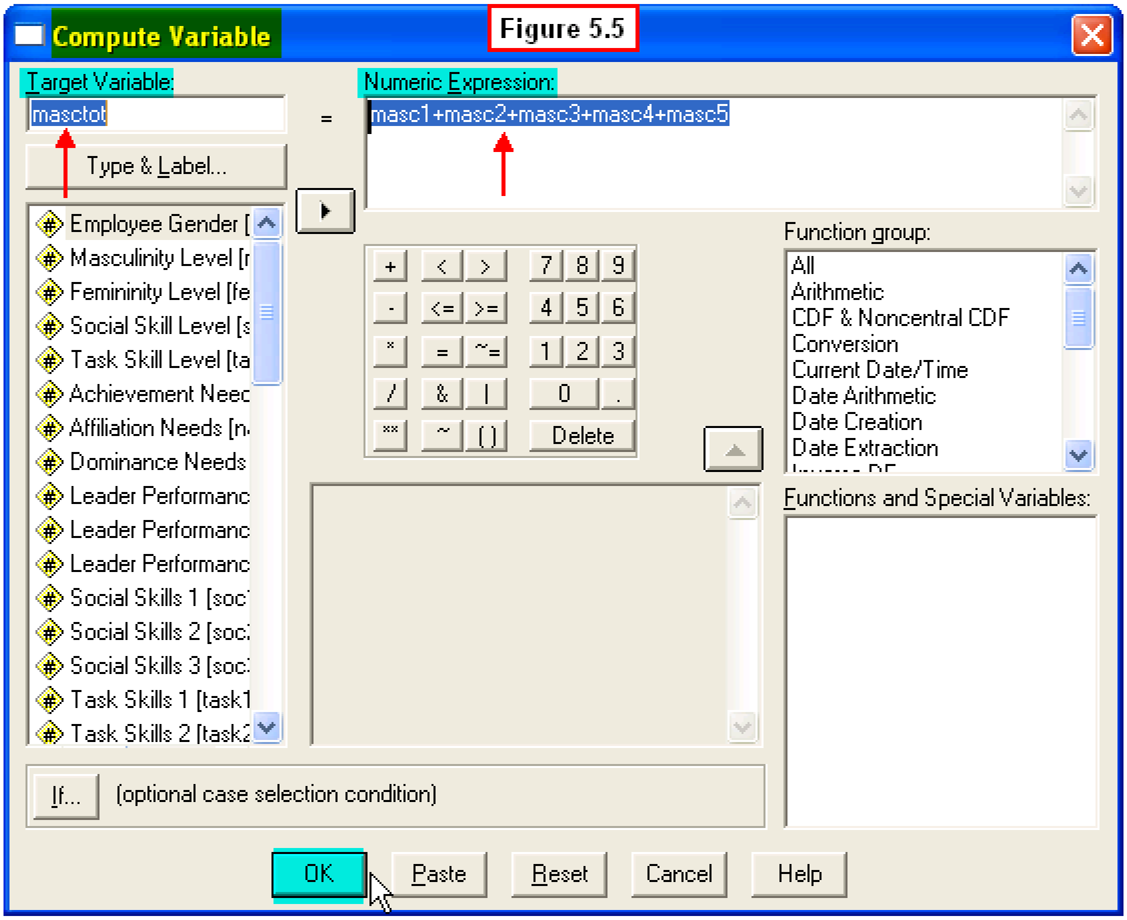
In this section, we explain how the new variables just described were created. All of this was done using an important SPSS feature called **Transform**. This feature allow the researcher to modify existing raw data into more useful forms for analysis.

5.2a The Compute Procedure

Recall that the **masc1-masc5** and **fem1-fem5** variables consist of self-ratings of how descriptive employees thought individual traits (stereotyped masculine and feminine) were of themselves. While some research questions might require the use of individual items such as these, more frequently researchers combine scores on individual items in a questionnaire into one total score. Thus, we combined these individual items into a single summated rating. This is done easily enough using the **Compute** subroutine of the **Transform** procedure. To compute a new variable by summing scores on several variables, simply select **Transform, Compute** from the Data Editor menu (see Figure 5.4).



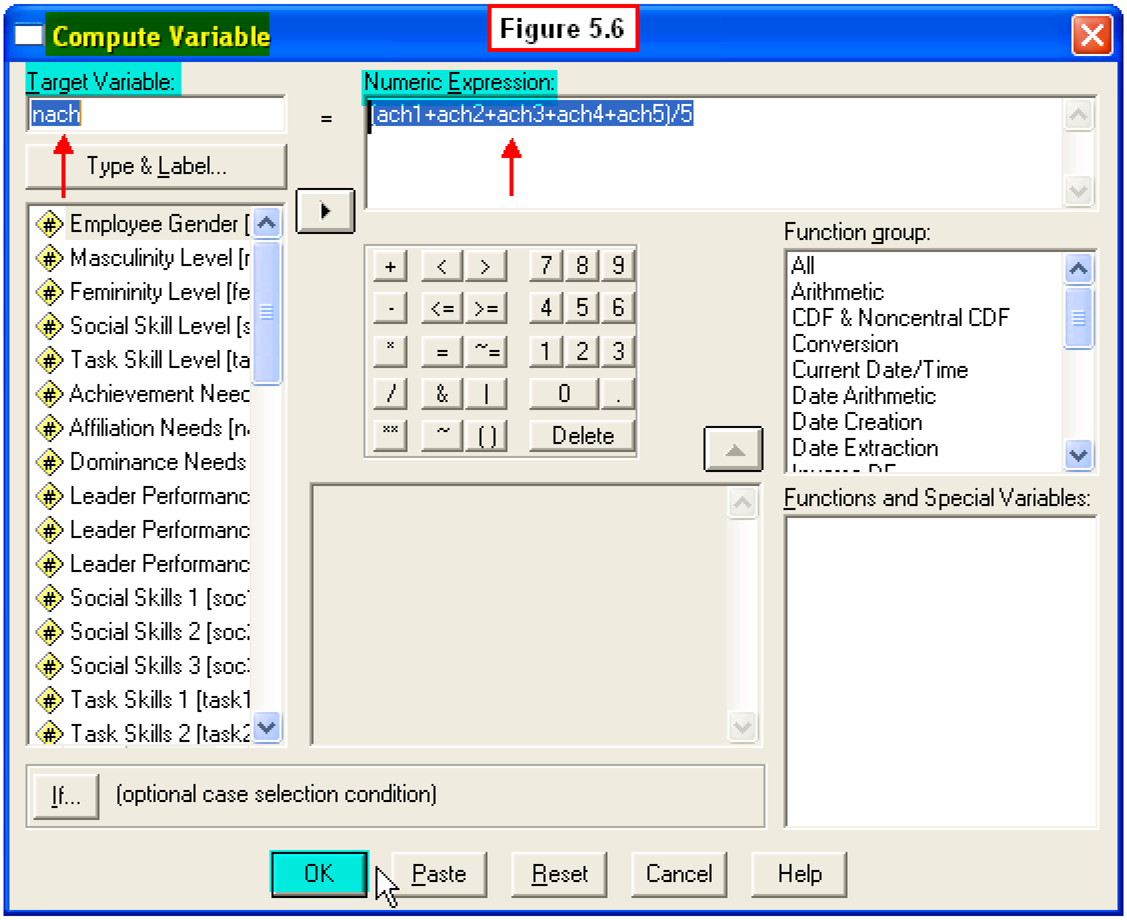
A **Compute Variable** dialog box will appear (see Figure 5.5). To name the first new variable to be created, we typed **masctot** in the **Target Variable:** box. To define the computation to be performed, we typed **masc1+masc2+masc3+masc4+masc5** in the **Numeric Expression** box. After we clicked **OK**, SPSS added the five masculine trait scores for every employee and put the total in a new column labeled **masctot** (our new variable).



As mentioned, SPSS tacks this new variable in the last column of the data file. Note that even though we used the **masc1-masc5** variables to generate the new **masctot** variable, the original variables are still retained in our file. This is convenient should we want to analyze the individual traits at some point. But more importantly, we have also gained a new variable which will be much easier to work with.

Show Me Video!

We followed this same procedure to compute **femtot**. We also used the **Compute** procedure to create the new variables, **nach**, **naff** and **ndom**. This time, however, we typed in the formula for computing the average of the five individual scores on the original **ach1-ach5**, **aff1-aff5** and **dom1-dom5** work motive variables. See Figure 5.6 for an example of this procedure used to compute **nach**.



Next we turn to another **Transform** procedure subroutine, called **Recode**. This procedure allows us to create a new variable by recoding scores on another variable. This is the procedure we used to convert the continuous variables, **masctot** and **femtot** into the categorical variables, **masc** and **fem**. We also did this to create the new variables, **soc** and **task**.

It is often desirable to transform **continuous variables** (variables whose scores are on a quantitative continuum, such as from 1 to 7) into **dichotomous variables** (variables whose scores represent two different categories). Researchers sometimes do this in order to be able to treat a continuous person variable (e.g., degree of masculinity, as measured in **masctot**) as a **quasi-experimental** independent variable consisting of two groups (e.g., low vs. high masculinity, as we have done with **masc**).

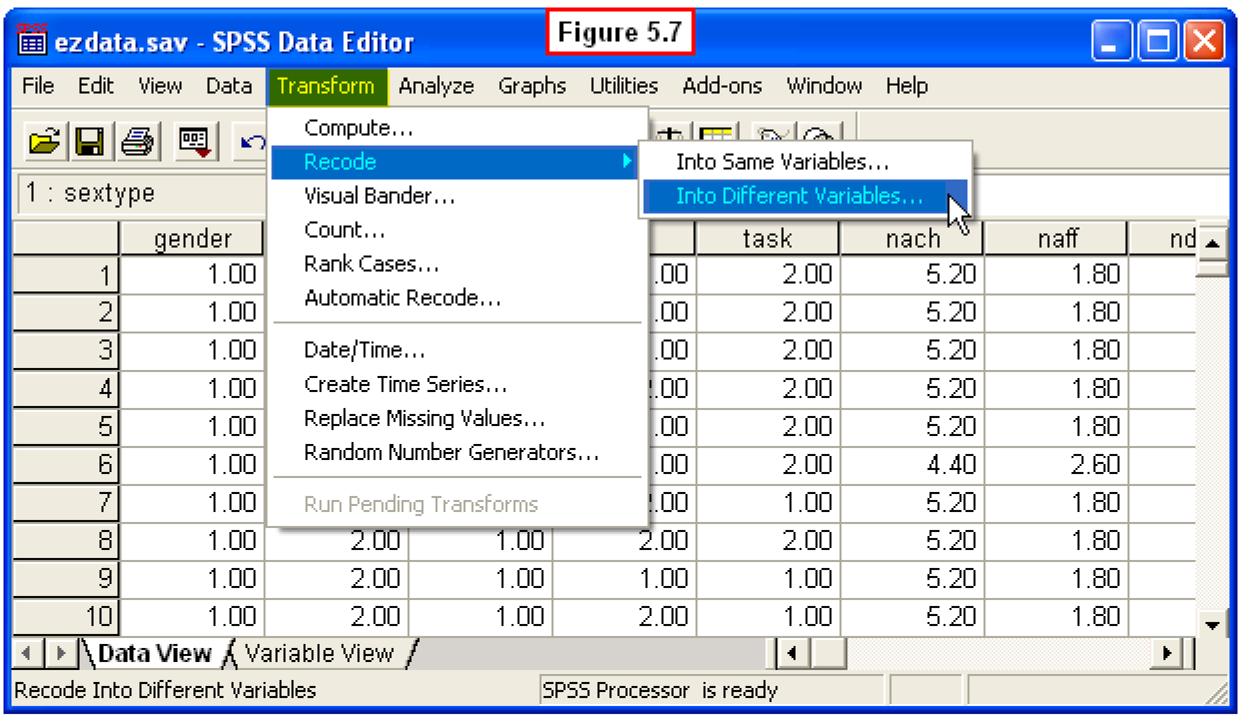
Thus, we transformed the scores on the continuous variable, **masctot**, (which range from 7 to 35) into a new dichotomous variable, **masc**, consisting of just two categories and two score values (1: **low masculinity**; 2: **high masculinity**). Once this is done, we can treat this new**masc** variable as an independent variable to examine differences in leader performance between employees in the high-masculinity category compared to those in the low-masculinity group. Although this may sound complicated, the process is relatively simple using the **Recode** procedure.

5.2 b The Recode Procedure

To create this new variable, we used the following recoding rules:

- If an employee's **masctot** score is **17 or lower**, then that score will be transformed to a value of **1** on the new **masc** variable (i.e., s/he will be placed in the **low masculinity** category)
- If an employee's **masctot** score is **18 or higher**, then that score will be transformed a value of **2** on the new **masc** variable (i.e., s/he will be placed in the **high masculinity** category).

To produce this transformation, we selected **Transform, Recode, Into Different Variables** from the menu in the Data Editor (Figure 5.7).



This caused a **Recode into Different Variables** dialog window to appear (Figure 5.8). We scrolled down the left pane and highlighted **masctot**, then clicked the right arrow button between panes to move this variable to the **Input Variable - > Output Variable:** pane.

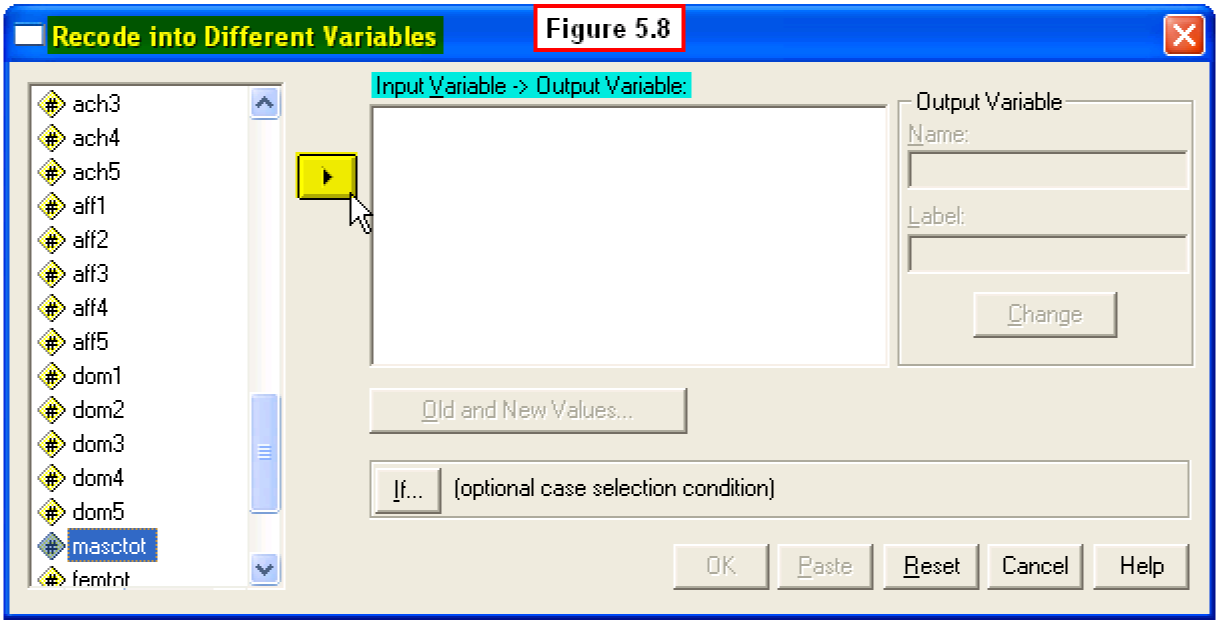


Figure 5.8

SPSS now expects the user to name the new output variable, as can be seen by the **masctot - > ?** question that appears in the **Numeric Variable - > Output Variable:** pane (Figure 5.9). We named the new output variable by typing **masc** into the **Output Variable** box on the right, then clicked the **Change** button to complete the naming.

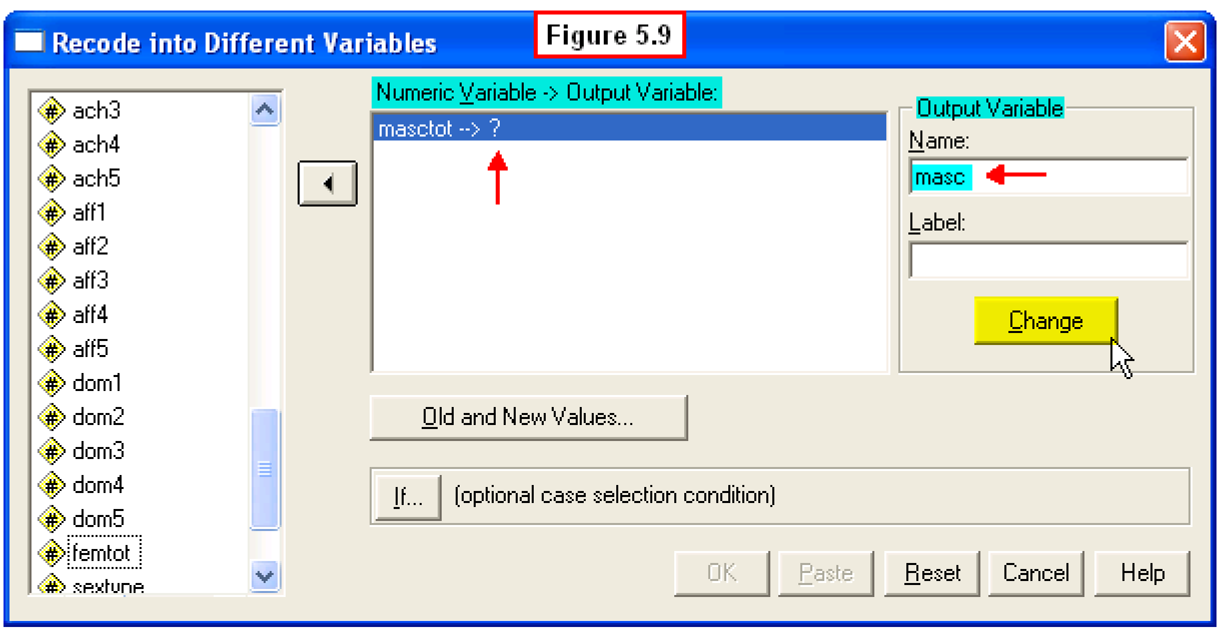
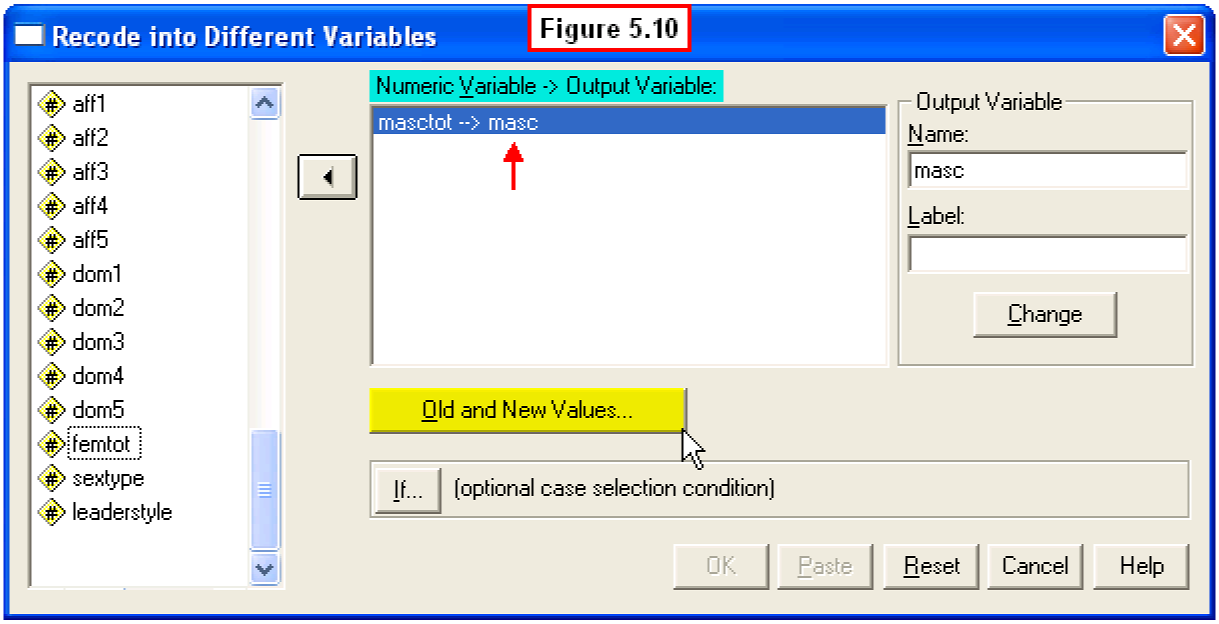
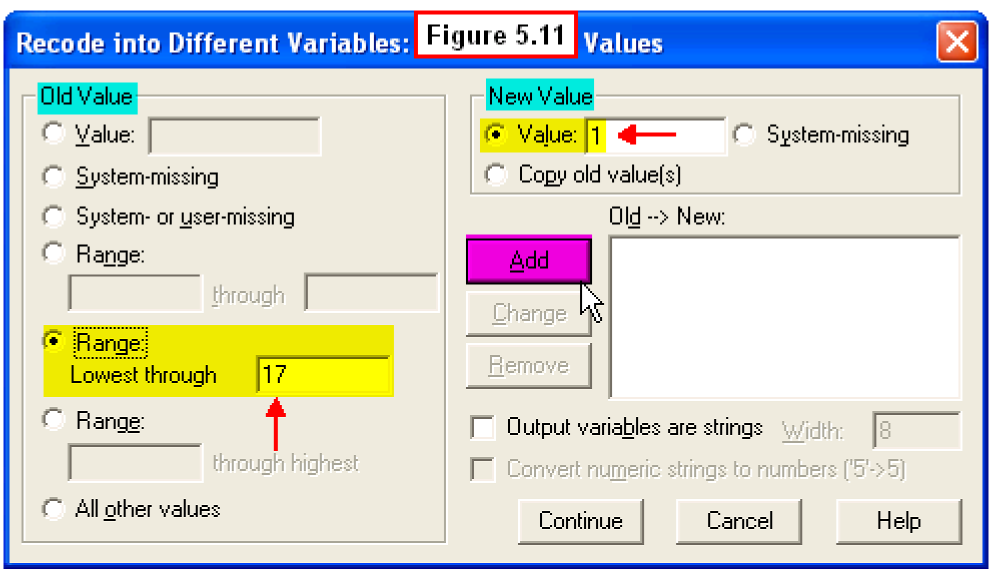


Figure 5.9

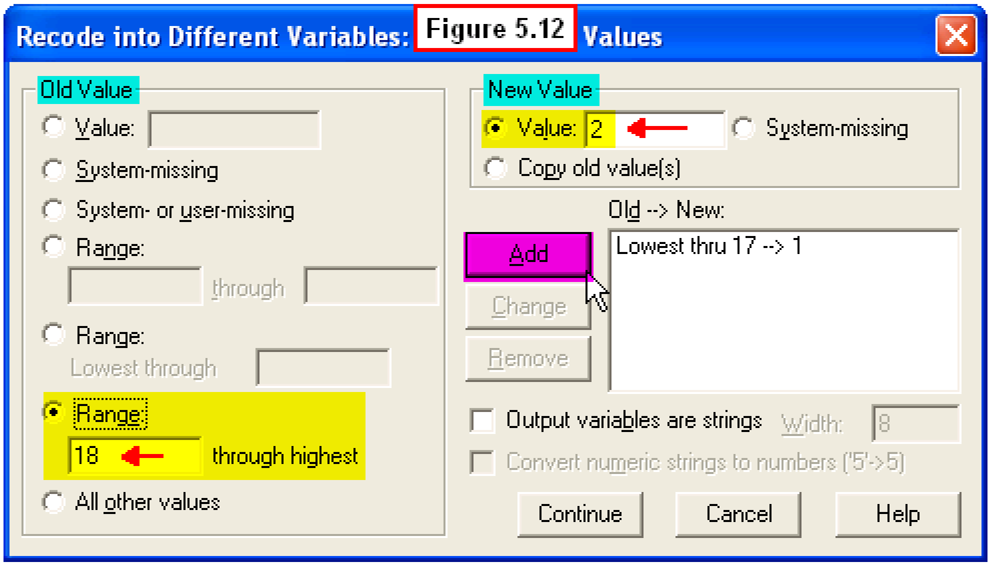
The new variable name (masc) was now displayed in the **Numeric Variable - > Output Variable:** pane. Next, to specify the recode rules to be used in creating the new output variable, we clicked the **Old and New Values** button (Figure 5.10).



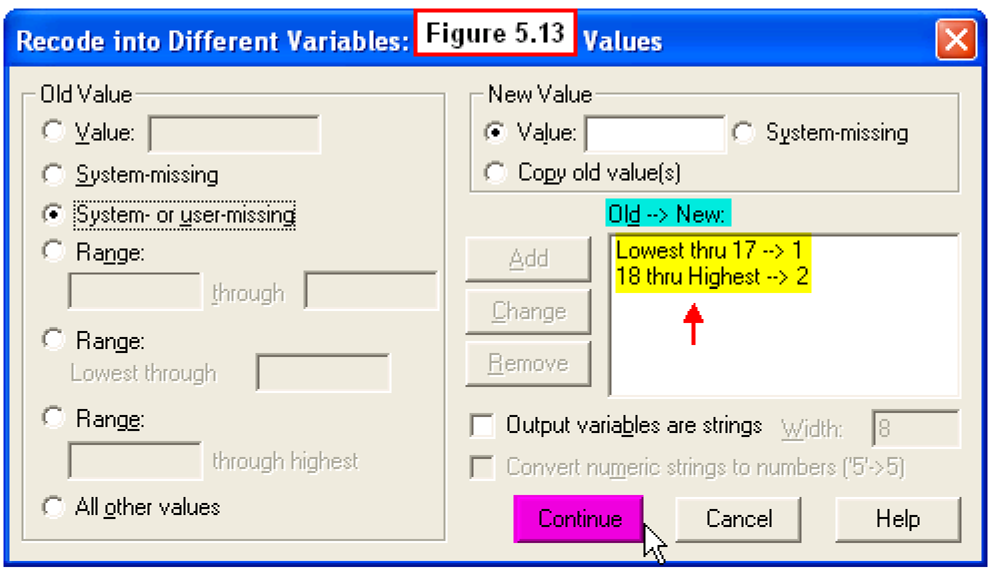
This caused yet another dialog window to appear (**Recode into Different Variables: Old and New Values**). As shown in Figure 5.11, we specified **Range Lowest through 17** in the **Old Value** pane, and specified the **New Value** as **1** on the right. This instructs SPSS to recode any employee's **masctot** score that is lower than 18 into the new value of 1 on the new**masc** variable (the **low-masculinity** category). Last, we clicked the **Add** button to complete this variable transformation.



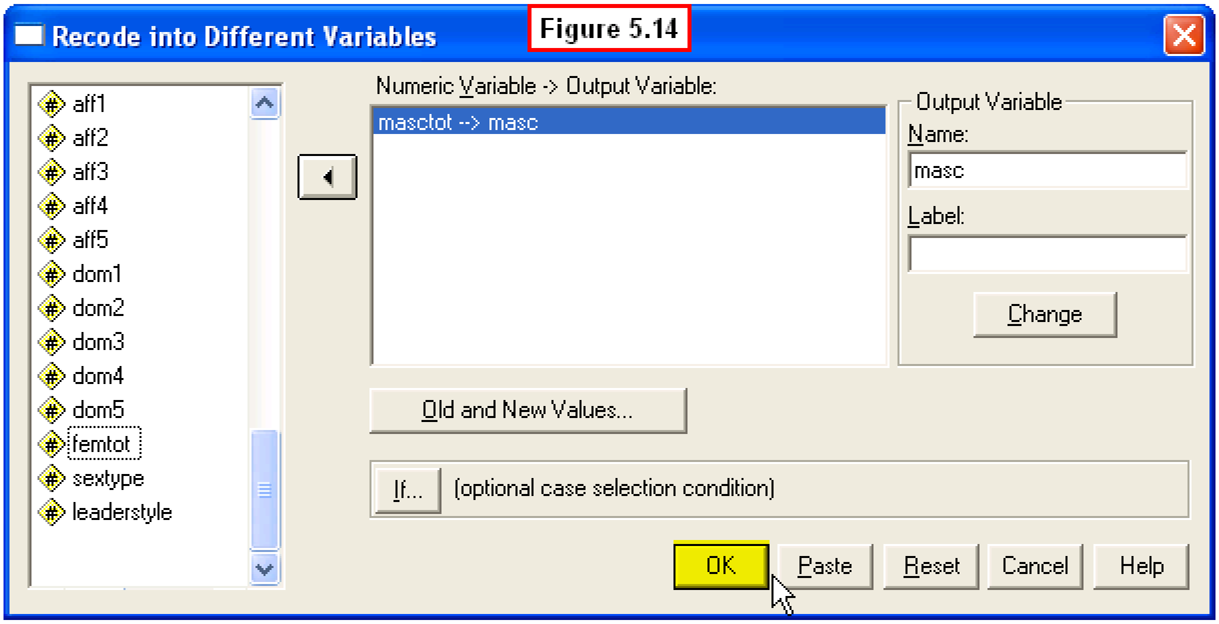
As shown in Figure 5.12, we then specified **Range: 18 through highest** in the **Old Value** pane, and specified the **New Value** as **2** on the right. This instructs SPSS to recode any employee's **masctot** score that is higher than 18 into the new value of 2 on the new **masc** variable (the **high-masculinity** category). Again we clicked the **Add** button to complete this variable transformation.



The rules we just specified are now summarized in the **Old -- > New:** pane on the lower right (Figure 5.13). Next, we clicked the **Continue** button at the bottom of this dialog window.



This caused the **Old and New Values** dialog window to close. For the last step, we clicked the **OK** button at the bottom of the **Recode into Different Variables:** dialog window (Figure 5.14).



SPSS then performed this transformation and added the new recoded **masc** variable at the end of the data file (recall that we then cut and pasted this new variable so that it now appears as the second variable in the **ezdata.sav** file you downloaded). If you scroll down your Data Editor window, you will see that all the values of this new variable are either **1** (low-masculinity) or **2** (high masculinity).

Show Me Video!

We used exactly the same **Recode** procedure to create the new **fem** variable, using the same recoding rules. That is:

- If the **femtot** score was **17 or less**, then the **fem** score = **1** (**Low-Femininity**)
- If the **femtot** score was **18 or higher**, then the **fem** score = **2** (**High-Femininity**)

Last, we used the **Recode** procedure to create the new categorical variables, **soc** and **task**, by recoding the original **soc1** and **task1** variables. Since soc1 and task1 are variables with score ranges from 1 to 9, we used a different set of recoding rules:

- If the **soc1** score was **4 or less**, then the **soc** score = **1** (**Low Social Skills**)
- If the **soc1** score was **5 or higher**, then the **soc** score = **2** (**High Social Skills**)
- If the **task1** score was **4 or less**, then the **task** score = **1** (**Low Task Skills**)
- If the **task1** score was **5 or higher**, then the **task** score = **2** (**High Task Skills**)

You're probably glad that we performed all of these transformations ourselves in the **ezdata.sav** file that you downloaded! It might also seem like a lot of variables to keep straight, but you will become more familiar with them once we begin doing data analyses involving these variables.

Further, although the **Compute** and **Recode** procedures we have described can seem confusing at first (and tedious to perform), we will give you a chance do some simpler transformations at the end of this chapter. Once you have tried them, you should see that the procedures themselves are fairly straightforward.

Before we do this, in the last section of this chapter we describe one more housekeeping procedure that makes the data file (and outputs from analyses) much easier to understand: providing more descriptive names for variables and adding labels to the levels of categorical variables.

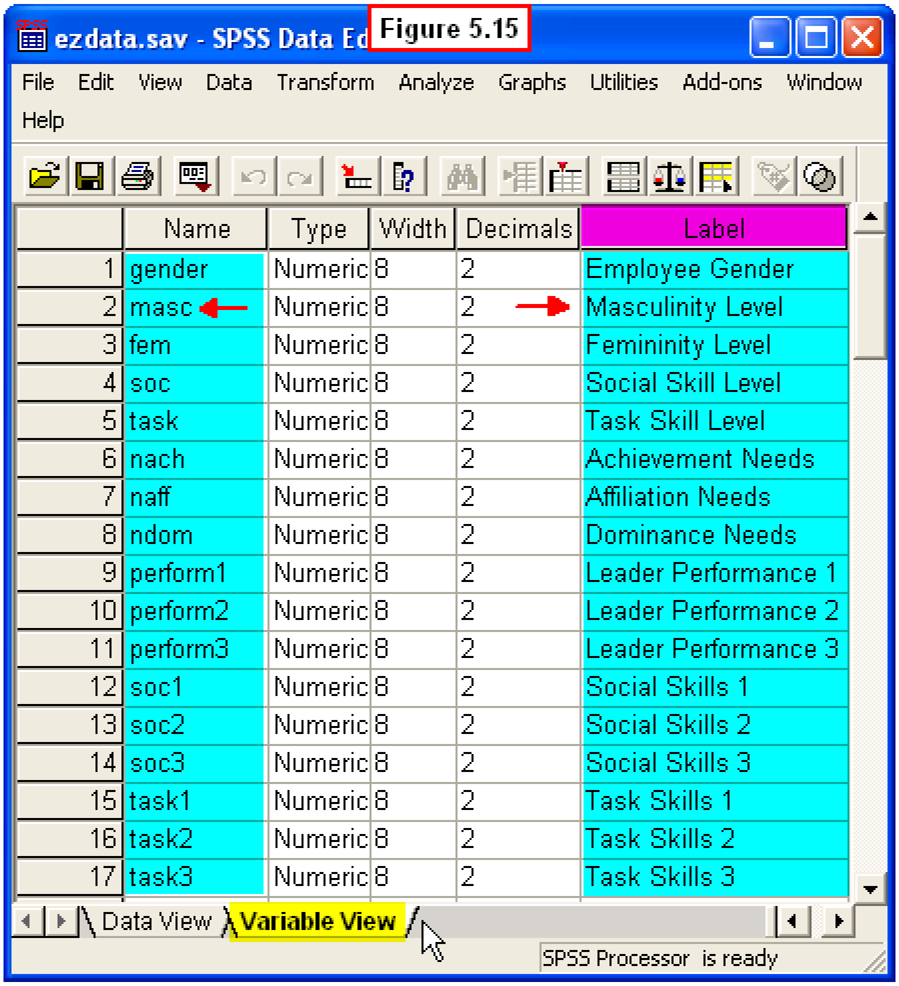
5.3 Adding Variable Labels & Value Labels

One helpful feature of SPSS that reduces confusion and information overload allows the user to provide more descriptive names (labels) for the variables in the file. Recall that SPSS has some restrictive rules for naming variables, which is why we have chosen such short abbreviations for our variable names.

But it can be hard (especially at first) to remember that, for example, the variable **masc** refers to the employee's level of masculinity. Further, it can be hard to remember that a value of 1 on **masc** means low-masculinity and a score of 2 means high masculinity. SPSS allows us to avoid this problem by adding variable and value labels to our data file.

5.3a Adding Variable Labels to the Data File

We have already added variable labels to the major variables in the **ezdata.sav** file that you downloaded. Click the **Variable View** tab of the Data Editor to see these labels (Figure 5.15).



To provide these descriptive variable labels, we simply clicked on the appropriate cell of the **Label** column and typed in the label we wanted to provide for a given variable. For example, we typed **Masculinity Level** as a more descriptive label for the **masc** variable. These labels will appear on output files in place of the shorter variable name. This will make interpreting the output files much easier when we conduct data analyses. Next we turn to the process of adding value labels to the categorical variables in our data file.

As you are probably well aware by now, it can be difficult to remember what the various numerical values assigned to levels of our categorical variables (**gender**, **masc**, **fem**, **soc** and **task**) mean. The Values feature of SPSS allows us to provide descriptive *verbal* labels for these *numerical* values. As with variable labels, value labels will appear in output files. Thus, you will not have to remember that a score of 2 for gender means a female employee - instead of 2, the word **Female** will be displayed in the output file.

5.3b Adding Value Labels to the Data File

This procedure is also accomplished from the **Variable View** in the Data Editor (Figure 5.16).

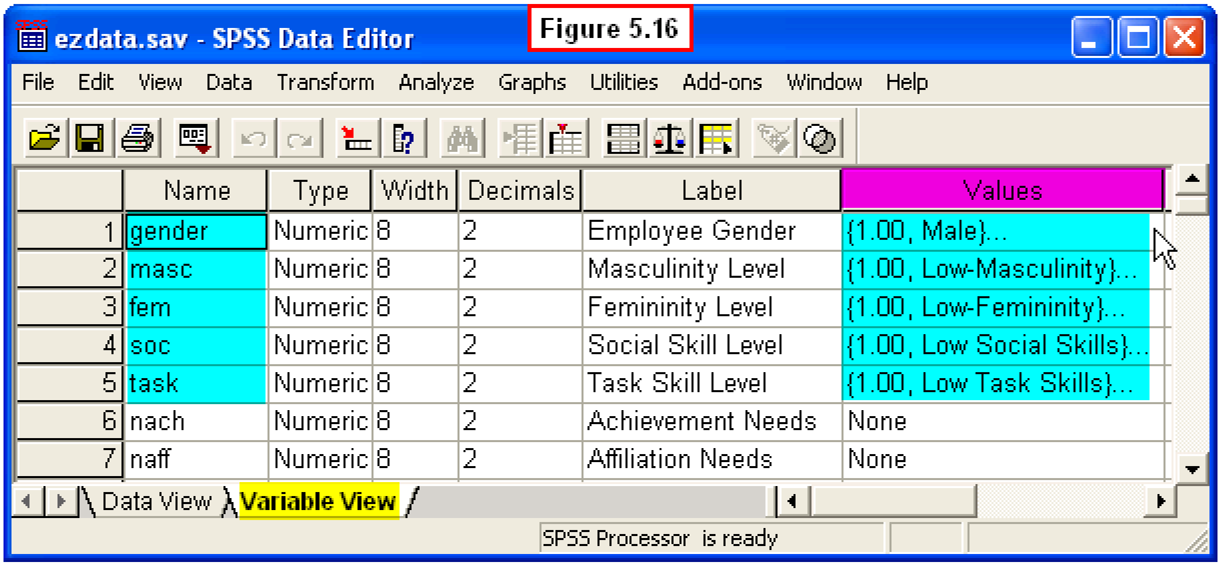


Figure 5.16

The first part of each value label we added for each variable is shown in the **Values** column of this Data Editor view. As an aside, you may be wondering why there are no value labels shown for other variables in the file (e.g., **nach**). The reason is that the other variables are continuous (as opposed to categorical). Since these types of variables can take on many different values (e.g., an average achievement need of 6.43), it would be unwieldy (and unhelpful) to attempt to provide a verbal label for every possible numerical value of continuous variables.

To see how we added these value labels, click on the first cell in the **Values** column (i.e., the value labels for **gender**). When you do this, a grey box with three dots (**...**) will appear to the right of the **(1.00, Male)...** that you see in this cell (Figure 5.17).

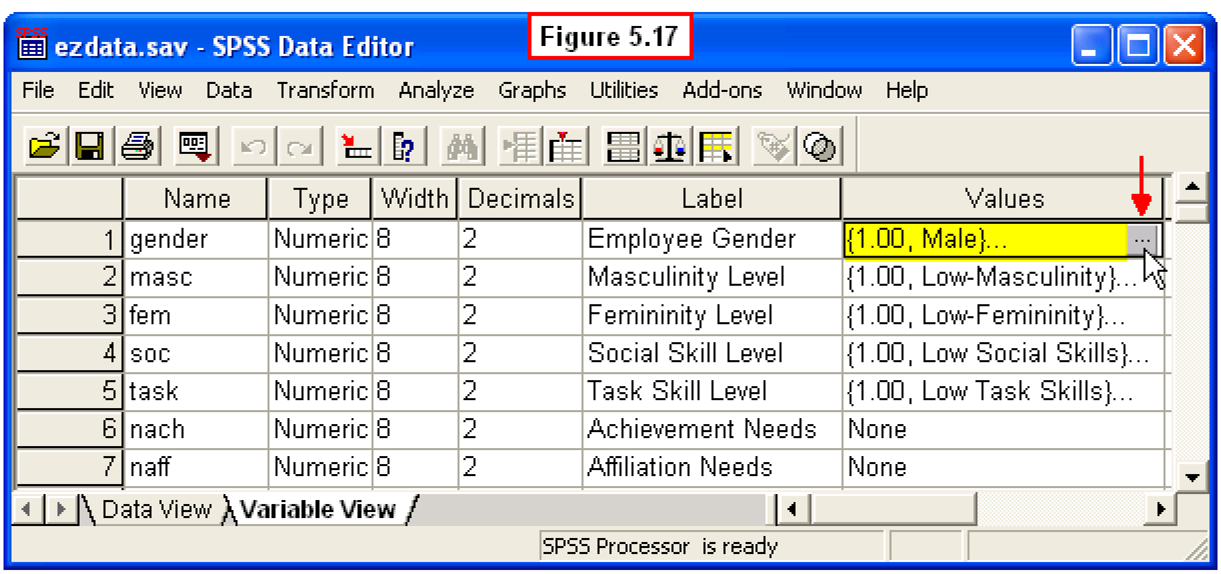
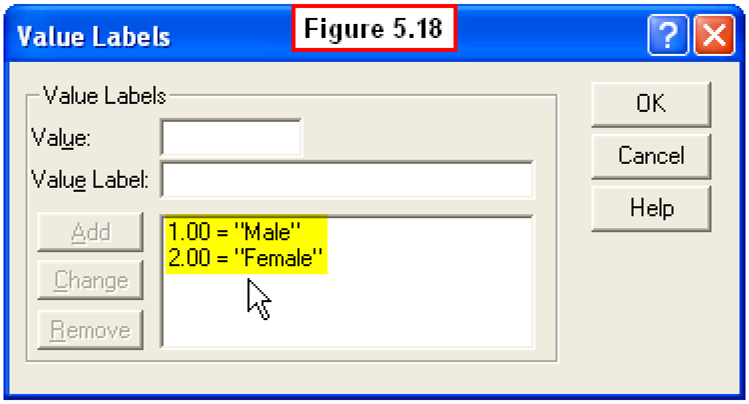
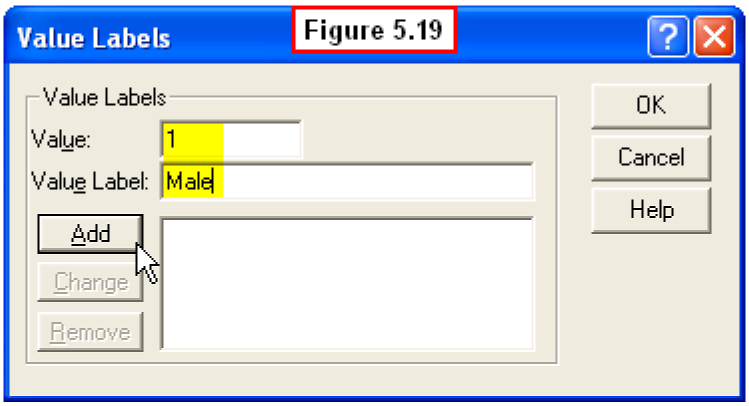


Figure 5.17

Click on the grey box and a **Value Labels** dialog window appears (Figure 5.18).



Note that since we have already provided the labels, they appear in the lower pane of Figure 5.18. In Figure 5.19 we illustrate how we created the first label (1.00 = "Male").



As can be seen, this simply involved typing **1** in the **Value:** box and typing **Male** in the **Value Label:** box, then clicking the **Add** button. To create the second value label, we repeated this procedure, typing **2** for value and **Female** for value label, then clicking the **Add** button again.

The same procedure was followed to create the value labels for the other four categorical variables (**masc**, **fem**, **soc** and **task**). You should click on the cells under the **Labels** column for these other variables to familiarize yourself with the labels we have added for them.

Show Me Video!

In Figure 5.20 we show an example of a frequency table we generated that displays the variable label and value labels for **gender.**

Figure 5.20

Variable Label

Employee Gender

Value Labels

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	110	48.2	48.2	48.2
	Female	118	51.8	51.8	100.0
	Total	228	100.0	100.0	

As this figure shows, the variable and value labels are well worth the time to create, since they greatly facilitate the ease of interpreting output of data analyses.

5.4 Chapter Review Video

[Review Me!]

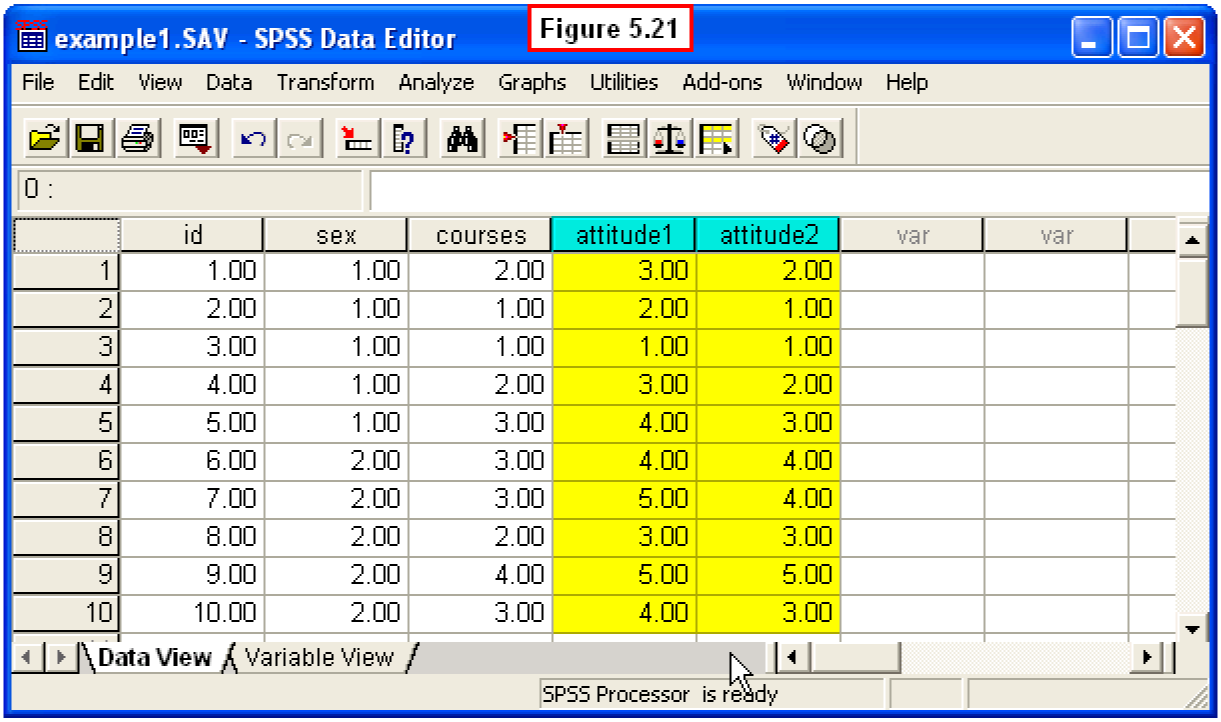
5.5 Try It! Exercises

1. Creating New Variables, Transforming Variables & Adding Verbal Labels

For these exercises, you will need to open the **example1.sav** file that you created in Chapter 2. You will be asked to type in scores on two new variables in this data file:

- student attitudes towards computer-computation of statistics (**attitude1**)
- student attitudes towards hand-computation of statistics (**attitude2**)

1. Enter the scores on the two new variables we have added to this file (Figure 5.21).

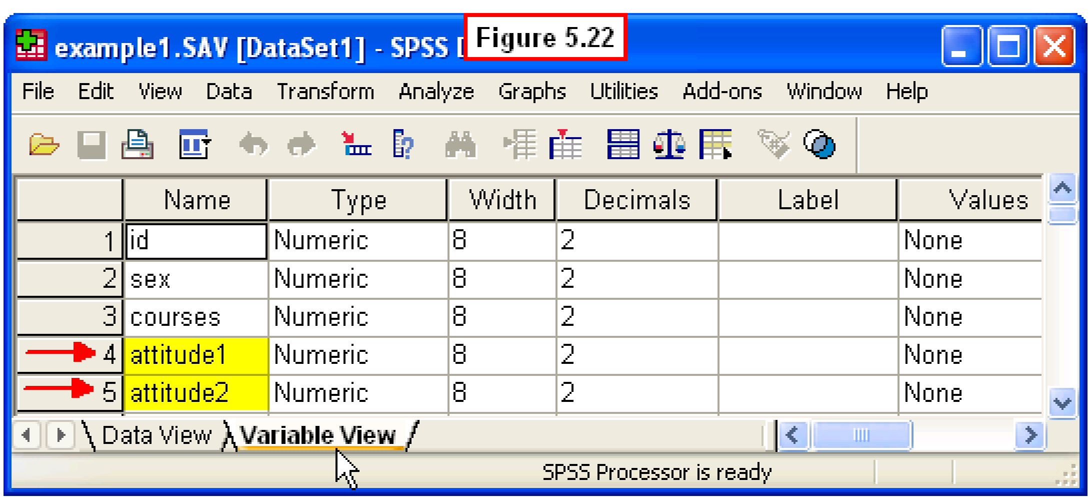


	id	sex	courses	attitude1	attitude2	var	var
1	1.00	1.00	2.00	3.00	2.00		
2	2.00	1.00	1.00	2.00	1.00		
3	3.00	1.00	1.00	1.00	1.00		
4	4.00	1.00	2.00	3.00	2.00		
5	5.00	1.00	3.00	4.00	3.00		
6	6.00	2.00	3.00	4.00	4.00		
7	7.00	2.00	3.00	5.00	4.00		
8	8.00	2.00	2.00	3.00	3.00		
9	9.00	2.00	4.00	5.00	5.00		
10	10.00	2.00	3.00	4.00	3.00		

Assume that these scores represent favorableness of the attitude for computer/hand computation:

- **attitude1**: Attitude towards computer computation (1 = Unfavorable; 5 = Favorable)
- **attitude2**: Attitudes towards hand computation (1 = Unfavorable; 5 = Favorable).

Note that before you enter the scores, you should first click the **Variable View** tab and type in the two variable names (Figure 5.22).



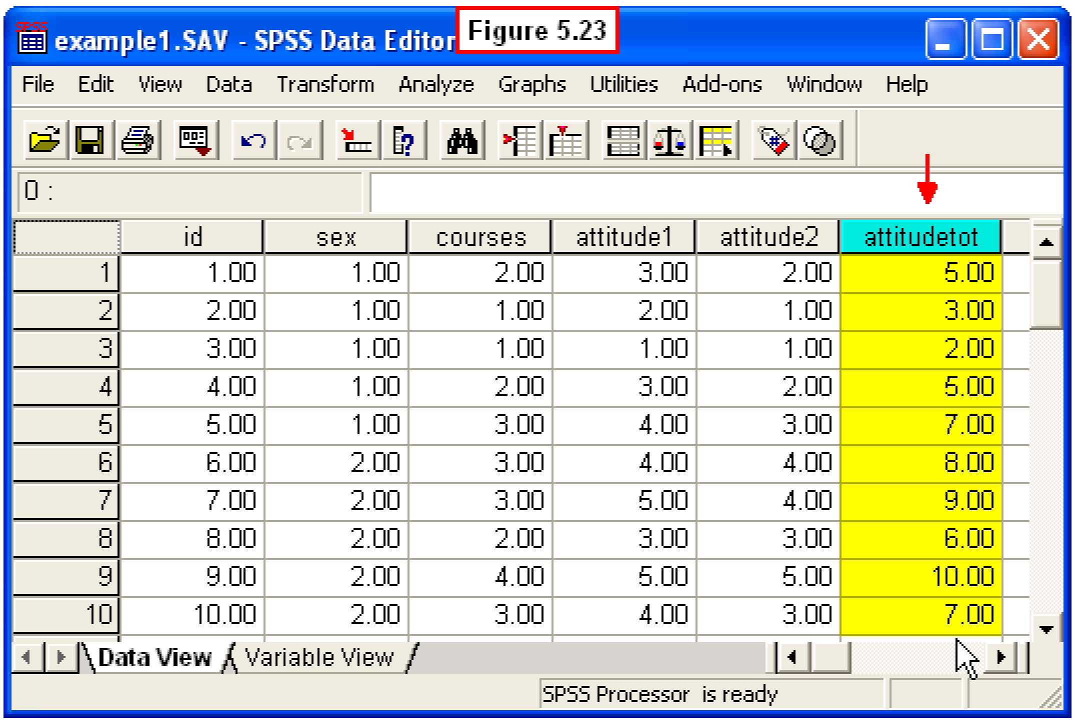
	Name	Type	Width	Decimals	Label	Values
1	id	Numeric	8	2		None
2	sex	Numeric	8	2		None
3	courses	Numeric	8	2		None
4	attitude1	Numeric	8	2		None
5	attitude2	Numeric	8	2		None

Then click the **Data View** tab and enter the scores for these two new variables shown in Figure 5.21.

2. Use the Compute Procedure explained in Section 5.2a to compute a new variable (**attitudetot**) which is the sum of the two new attitude scores:

- Select **Transform, Compute** from the Data Editor menu
- Type **attitudetot** in the **Target Variable:** box of the Compute dialog window
- Type **attitude1+attitude2** in the **Numeric Expression:** box
- Click the **OK** button

When you are finished, the Data Editor should display the new **attitudetot** variable (Figure 5.23).



	id	sex	courses	attitude1	attitude2	attitudetot
1	1.00	1.00	2.00	3.00	2.00	5.00
2	2.00	1.00	1.00	2.00	1.00	3.00
3	3.00	1.00	1.00	1.00	1.00	2.00
4	4.00	1.00	2.00	3.00	2.00	5.00
5	5.00	1.00	3.00	4.00	3.00	7.00
6	6.00	2.00	3.00	4.00	4.00	8.00
7	7.00	2.00	3.00	5.00	4.00	9.00
8	8.00	2.00	2.00	3.00	3.00	6.00
9	9.00	2.00	4.00	5.00	5.00	10.00
10	10.00	2.00	3.00	4.00	3.00	7.00

3. Use the Recode Procedure explained in Section 5.2b to transform the continuous variable, **courses**, into a new dichotomous variable, **experience**, with two categories. Use these recoding rules:

- If the **courses** score is **2 or lower**, then **experience** = **1** (Low Experience)
- If the **courses** score is **3 or higher**, then **experience** = **2** (High Experience)

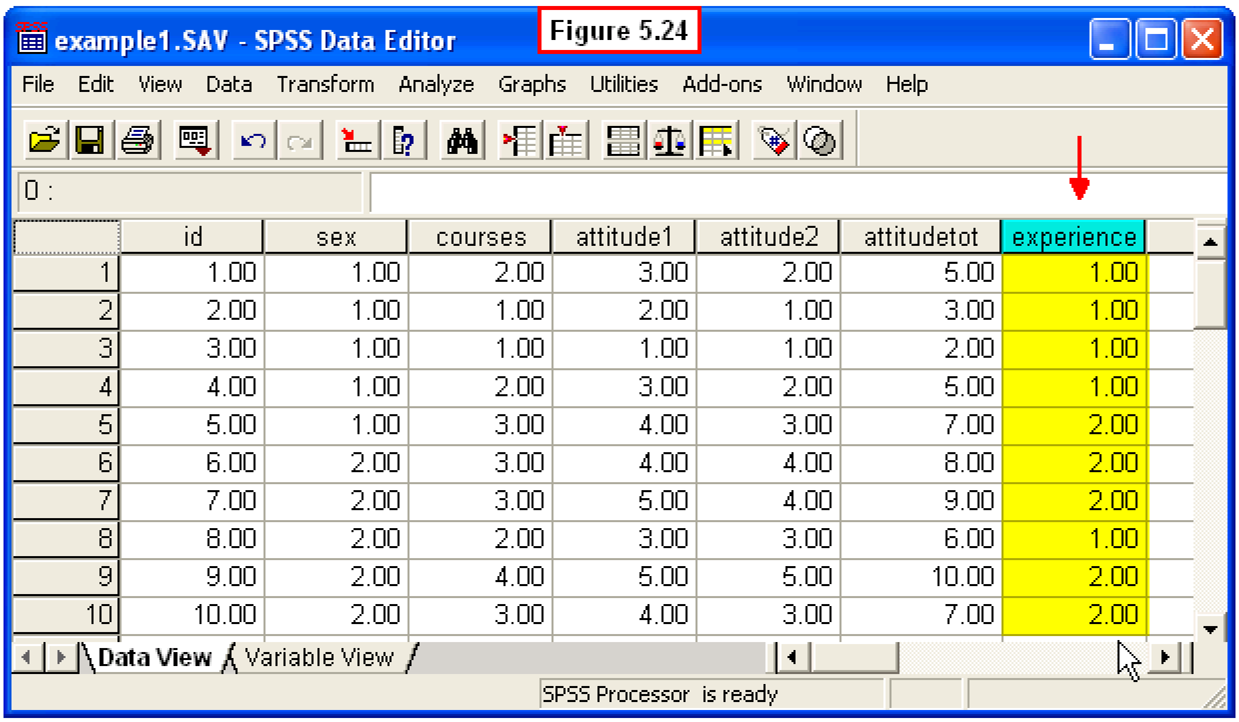
To perform this transformation, select **Transform, Recode, Into Different Variables** from the Data Editor menu. A **Recode into Different Variables** dialog window will appear. Do the following:

- Move the **courses** variable into the **Input Variable - > Output Variable** pane
- Type **experience** in the **Name:** box for the **Output Variable**
- Click the **Change** button, then click the **Old and New Values** button.

The **Old and New Values** dialog window will appear. Do the following:

- In the **Old Value** panel, select the **Range lowest thru___**button, then type **2** in the box
- In the **New Value** panel, type **1** in the box, then click the **Add** button
- Return to the **Old Value** panel, select the **Range____thru highest** button, then type **3** in the box
- In the **New Value** panel, type **2** in the box, then click the **Add** button, and then click the **Continue** button.

Last, click the **OK** button in the **Recode into Different Variables** dialog window. When you are done the Data Editor window should show the new categorical **experience** variable with scores of either 1 or 2 (Figure 5.24).



4. Add a descriptive Variable Label and Value Labels for the new **experience** variable using the procedures explained in Section 5.3b. First click the **Variable View** tab in the Data Editor.

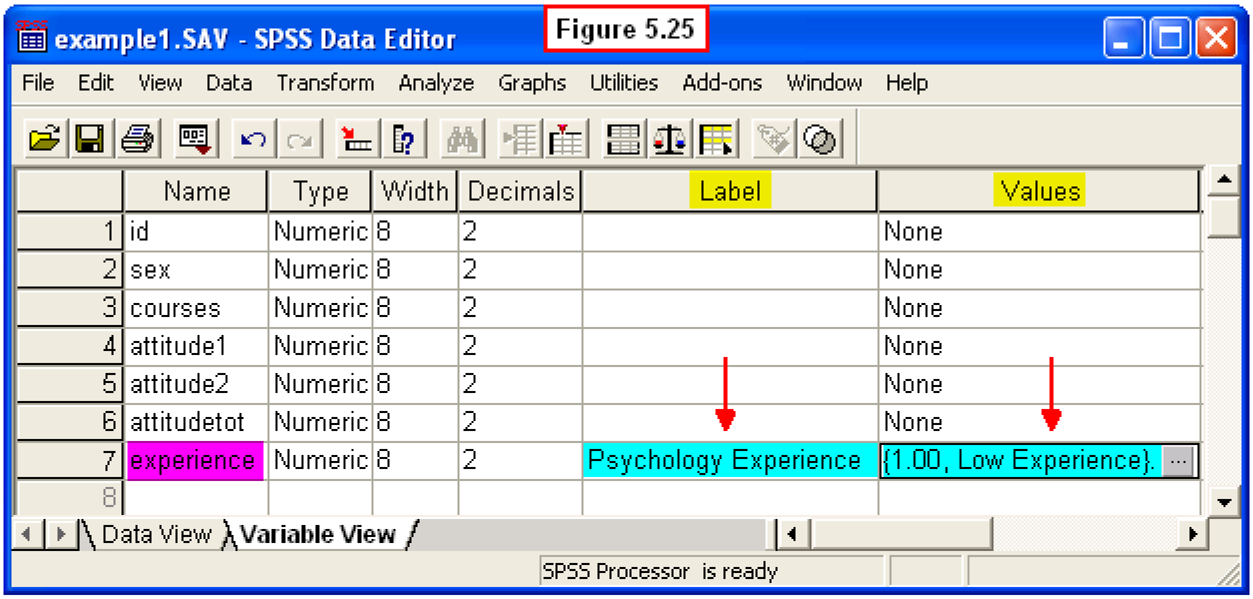
- Click on the cell in the **Label** column for the **experience** variable
- Type **Psychology Experience** in this cell.
- Click the cell for this variable in the **Values** column; then click the 3 grey dots (**...**)

A **Value Labels** dialog window will appear. To add the labels:

- Type **1** in the **Value:** box; type **Low Experience** in the **Value Label:** box; click the **Add** button

- Type **2** in the **Value:** box; type **High Experience** in the **Value Label:** box; click the **Add** button
- Click the **OK** button of the **Value Labels** dialog box

When you are finished, the Data Editor window should look like the one in Figure 5.25



5. Click the **Data View** tab, then save and **print** this file to submit to your instructor.

Chapter 6
The Frequencies Procedure: Summarizing Data With Descriptive Statistics

6.1 Introduction

One of the primary reasons for doing research is to be able to make accurate statements regarding the behavior or characteristics of a large number of people. In most cases it is impossible to actually collect data from every member of a target group (referred to as a **population**), so researchers typically collect data from a smaller subset of the population (called a **sample**) and attempt to generalize from the sample to the larger population. For example, 228 participants were selected from the several thousand employees of EZ Manufacturing for the leadership study.

But even when working with relatively smaller samples, it is very difficult for the human mind to comprehend large numbers of individual facts. Few researchers could remember all the individual performance scores of the 228 EZ Manufacturing employees, and it would be difficult to draw meaningful conclusions from this raw data itself. For this reason researchers often begin data analyses by summarizing characteristics of the participants in the sample.

One way to do this is to generate frequency tables of variables. Recall from Chapter 3 that frequency distributions summarize data by listing the number of participants who received scores of the possible values on the variables. Thus, a frequency table of how many employees scored a 1, 2, 3, etc. on the **perform** variable will be much easier to understand and interpret than would a simple listing of all 228 individual performance scores. This process is sometimes called *number crunching*, because large volumes of data are crunched into more manageable, meaningful units. This facilitates drawing conclusions and determining trends in the data (e.g., do most of the employees score towards the high or low end of the performance scale?).

Beyond constructing simple frequency distributions, researchers typically crunch data even further down into a single statistic that is typical of the entire set of scores in some way. These numerical indices are called **descriptive statistics**, because they provide a single number, or index, that best summarizes all of the scores on some dimension. Common descriptive statistics employed are measures of central tendency and variability.

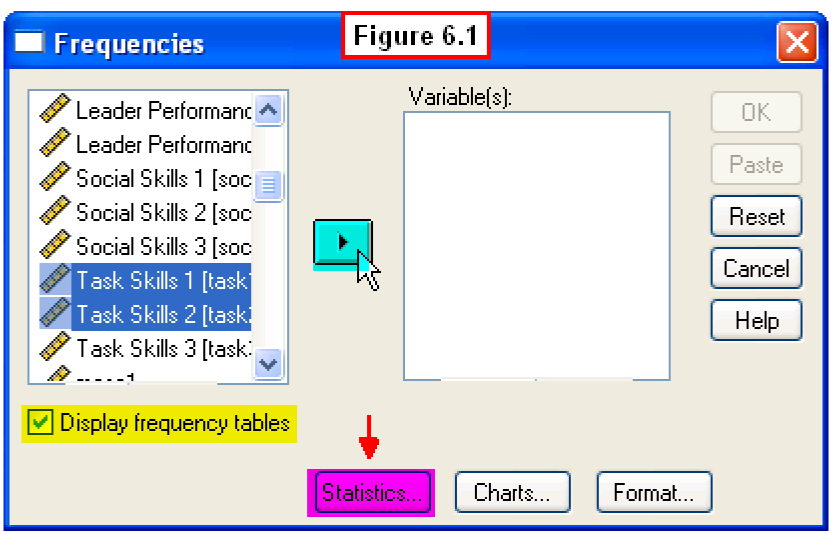
The mean, median, mode, are common measures of central tendency (so called because these indices tend toward the middle, or center, of the distribution). The standard deviation, variance, and range are examples of indices of variability (so called because they describe the how much all the scores vary around the middle of the distribution). These summary statistics go a long towards helping the researcher understand the data, and they enable her/him to describe the findings in a brief, precise manner.

Recall that we used the **Frequencies** procedure in Chapter 3 as an introduction to SPSS. In this chapter we will use this procedure to generate frequency tables of two variables in the **ezdata.sav** file. We will also demonstrate how this procedure can be used to generate and interpret descriptive statistics for these variables. Open your **ezdata.sav** file and follow along with the example. You will be asked to do the same analyses on different variables in the file as an exercise at the end of the chapter.

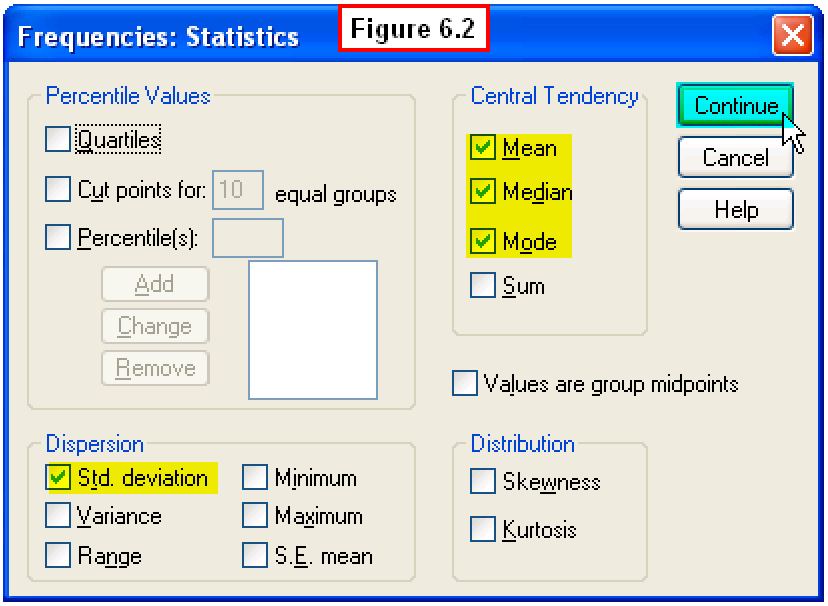
6.2 Running the Frequencies Procedure

Recall our Chapter 4 discussion of social and task skills in leadership. It would be of interest to describe EZ employees on these dimensions. Also recall that you obtained scores on these skills before and after employees attended a leadership training workshop. So it would also be of interest to compare the distribution of employees' scores on these skills before and after the workshop to begin to assess its effectiveness. For this example we will generate descriptive statistics for the before-after scores on task skills, **task1** and **task2**. To begin, select **Analyze, Descriptive Statistics, Frequencies...** from the Data Editor menu.

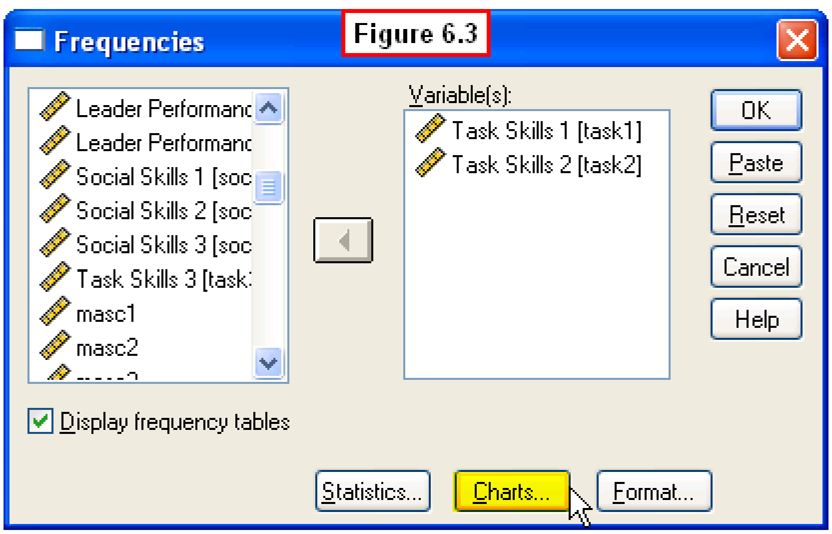
In the left pane of the **Frequencies** dialog window (Figure 6.1), scroll down and highlight **task1** and **task2**. Click the right-arrow in the middle to move these to the **Variable(s):** pane. Make sure the **Display frequency tables** box is checked, then click the **Statistics...** button.



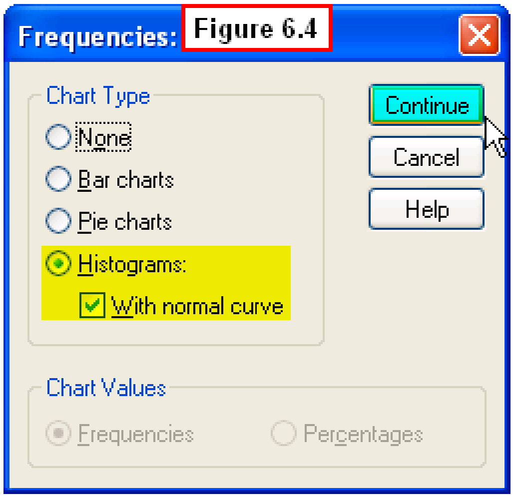
In the **Frequencies: Statistics** window (Figure 6.2), click the Mean, Median and Mode for measures of Central Tendency and Standard Deviation for measures of Dispersion (variability), then click the **Continue** button.



This will return you to the **Frequencies** window. To generate frequency graphs of these variables, click the **Charts...** button at the bottom (Figure 6.3).



In the **Frequencies: Charts** window (Figure 6.4), select **Histograms, with normal curve** (to superimpose a normal curve over the histogram), then click the **Continue** button.



This will return you again to the **Frequencies** dialog window. Now simply click the **OK** button in the upper-right corner of that window.

To view a video of this procedure, visit http://youtu.be/vGypnHFveTU

6.3 Interpreting the Output

The descriptive statistics we requested are displayed in the first table of the output file (Figure 6.5). The mean, median and mode of **Task Skills 1** are 5.12, 5.00 and 5.00, respectively. These three measures of central value are in close agreement, indicating that the middle of the distribution happens to be the midpoint of our 9-point scale of task skills. Thus, the typical employee was about average on task skills at the beginning of the study. The standard deviation (1.90) indicates there is variability in task skills (on average, about 2 points from the mean).

Note that after employees attended the leadership training workshop, the measures of central tendency indicate that there was a subsequent overall increase in the mean, median and mode on **Task Skills 2** (5.50, 6.00 and 6.00, respectively). Determining whether or not this is a statistically significant increase in task skills requires use of inferential statistics (a topic we will discuss in a later chapter). However, this comparison of **task1** to **task2** does provide some initial evidence that the workshop was effective, in that the typical employee now has a task skills score one point above the midpoint of the task skills scale. The standard deviation of **task2** (1.94) indicates that there wasn't much change in the variability of **task2** scores compared to that of **task1**.

Figure 6.5

Statistics

		Task Skills 1	Task Skills 2
N	Valid	228	228
	Missing	0	0
Mean		5.1228	5.5000
Median		5.0000	6.0000
Mode		5.00	6.00
Std. Deviation		1.89895	1.94075

Examination of the frequency table of **task1** (Figure 6.6) confirms that the highest frequency of scores was 5 (53 employees, or 23.2%, received this score), and the Cumulative Percent column shows that 59.6% of employees received a score of 5 or lower (and 40.4% scored 6 or higher). We get a sense of the variability in scores by examining the other frequencies/percentages. The scores are fairly evenly distributed within 2 points of this middle, with few scores at either the extreme high or low end of the scale.

Figure 6.6

Task Skills 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	9	3.9	3.9	3.9
	2.00	6	2.6	2.6	6.6
	3.00	34	14.9	14.9	21.5
	4.00	34	14.9	14.9	36.4
	5.00	53	23.2	23.2	59.6
	6.00	34	14.9	14.9	74.6
	7.00	33	14.5	14.5	89.0
	8.00	16	7.0	7.0	96.1
	9.00	9	3.9	3.9	100.0
	Total	228	100.0	100.0	

Examination of the frequency table of **task2** (Figure 6.7) confirms that the highest frequency of scores increased to a value of 6 (55 employees, or 24.1%, received this score), and the Cumulative Percent column shows that now only 42.1% of employees received a score of 5 or lower (and now 57.9% scored 6 or higher). Examining the other frequencies/percentages, we see that the scores are fairly evenly distributed within 2 points of this middle, with fewer scores at either the extreme high or low end of the scale. However, there appeared to be a substantial increase in scores at the high end of the scale compared to **task1**. That is, 17.1% have scores of 8 or higher on **task2**, compared to 10.9% with 8 or higher on **task1**. This is additional evidence of the effectiveness of the leadership training workshop.

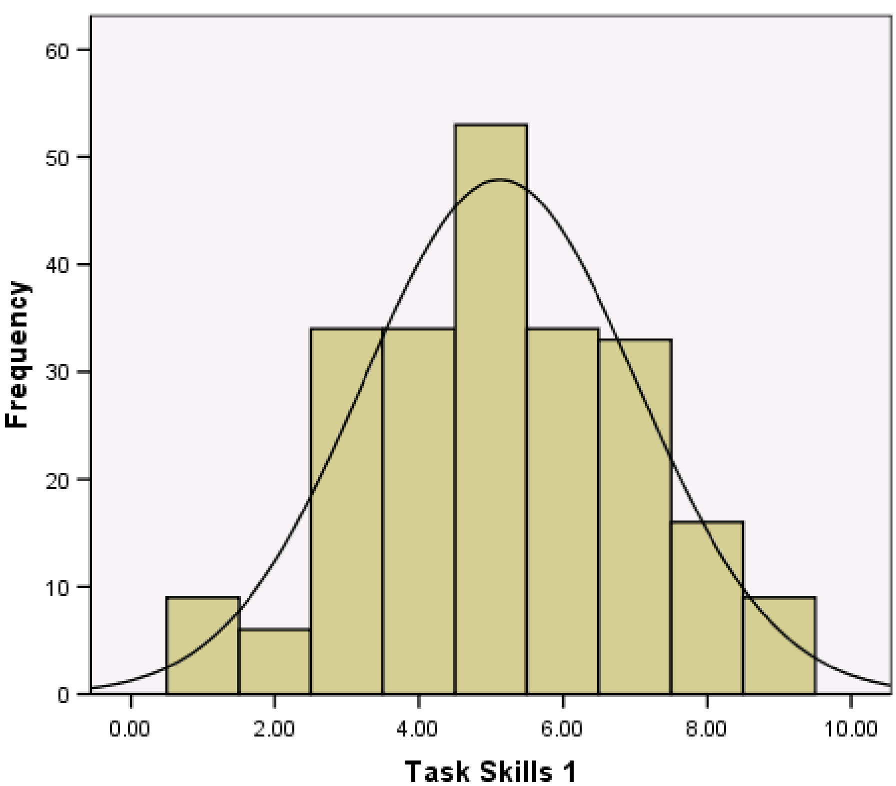
Figure 6.7

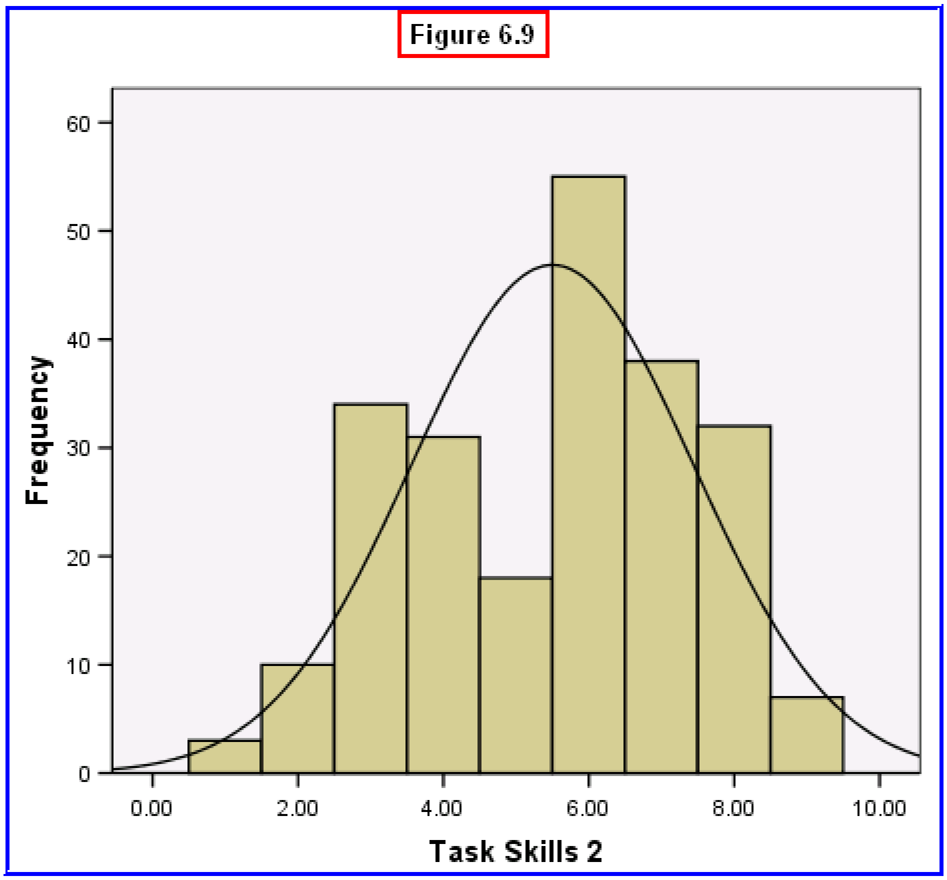
Task Skills 2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	3	1.3	1.3	1.3
	2.00	10	4.4	4.4	5.7
	3.00	34	14.9	14.9	20.6
	4.00	31	13.6	13.6	34.2
	5.00	18	7.9	7.9	42.1
	6.00	55	24.1	24.1	66.2
	7.00	38	16.7	16.7	82.9
	8.00	32	14.0	14.0	96.9
	9.00	7	3.1	3.1	100.0
	Total	228	100.0	100.0	

The histograms we generated visually depict this shift towards the higher end of the task skills scale (Figures 6.8 and 6.9).

Figure 6.8





Examination of these graphs reveals that the distribution has shifted to the higher end of the scale on **task2** scores. The bars depicting frequencies of scores 6 or higher are much taller in the second figure than in the first, reflecting an increased number of employees with high scores on task skills after attending the workshop.

More could be said about these results, but this has shown the value of the **Frequencies** procedure and computing **descriptive statistics** in helping the researcher gain an understanding of the data from a project such as this. This procedure is often the first step in analyses, and can be done for most variables in the data file. It can lead to some preliminary conclusions early on in the analysis phase, and it is sometimes used for exploratory purposes to generate suggestions for subsequent analyses (e.g., determining whether or not the increase in mean task skills scores is a statistically significant one).

To view a video review of this chapter's procedures, visit http://youtu.be/vGypnHFveTU

Chapter 7

Crosstabulation:
Understanding Bivariate Relationships Between Categorical Variables

7.1 Introduction to Crosstabs

So far we have discussed distributions of scores on a single variable. We now turn our attention to the topic of **bivariate distributions** - ones which characterize *relationships* between *two variables* simultaneously.

7.1a Bivariate frequency tables

These types of analyses examine variability in two or more distributions of scores to determine whether or not there is any pattern of covariation, or commonality, between the variables. In Chapter 8 we will discuss bivariate relationships between *quantitative* (continuous) variables. In the present chapter we consider relationships between *qualitative* (categorical) variables.

Much research involves data that are categorical rather than continuous. In fact, the variable transformations we described in Chapter 5 involved converting continuous variables (e.g. **masctot**, on which employees could score anywhere between 5 and 35) into dichotomous categorical variables (e.g., **masc**, in which employees were placed into just two groups - high vs. low masculinity).

We can examine relationships between categorical variables via an extension of frequency analysis. We could use the **Frequencies** procedure to construct a frequency table of the number of EZ employees who fall into the low vs. the high masculinity categories. However, a more interesting (and potentially more important) question we could ask is whether the frequency distribution of low vs. high masculinity is similar or different for male vs. female employees.

We could generate separate frequency tables of this variable for each sex to address this question, but that would be unsystematic. Further, SPSS includes a procedure that is specifically designed to generate frequency distributions for two variables simultaneously. This procedure is called **Crosstabs**, and it produces tables referred to as *crosstabulations*, because they tabulate or count the frequencies of values across two variables simultaneously.

Thus, **Crosstabs** allows us to answer questions such as whether there is a *relationship* between masculinity level and gender. A reasonable hypothesis, for example, might be that male employees are more likely than female employees to be in the high masculinity category, and female employees are more likely than male employees to be in the low masculinity category.

We could test this hypothesis by examining the frequencies obtained in a crosstabulation of the variables **masc** and **gender**. Our printout would contain a table that indicates:

- the number of men who are in the low-masculinity category
- the number of men who are in the high-masculinity category
- the number of women in the low-masculinity category
- the number of women in the high-masculinity

This will allow us to compare the number of men and women in the low/high masculinity categories, but it will also allow us to examine frequencies **within** gender. For example, the table would also permit us to examine whether women are more likely to be in the low-masculinity category than in the high-masculinity category, and vice-versa for men. Further, **Crosstabs** can calculate a statistic that allows us to determine whether any observed relationship between these variables is a statistically significant one or due to chance.

There are numerous other questions relevant to our project that could be addressed using **Crosstabs** to generate bivariate frequency distributions (for example, whether the high-masculinity category is most frequently associated with the high-task skill category of leader style). You're encouraged to explore these relationships (recall that we have several categorical variables in the **ezdata.sav** file: gender, masc, fem, task, soc). We will use the**mes** and **gender** example to illustrate the **Crosstabs** procedure in this chapter, and you will be asked to do the same analysis for **fem** and **gender** in the exercise at the end of the chapter.

7.1b The logic of the Chi Square statistic

The preceding discussion suggests that the two variables, **gender** and **masc**, are not independent, i.e., there is a *relationship* between one's gender and one's masculinity level. We will test this hypothesis by examining the frequencies obtained in a crosstabulation of the variables.

To illustrate how we could determine whether or not two variables are related using a crosstabulation table, we will simplify things for the moment and assume that there are only 50 men and 50 women in our sample who have been classified as either low or high in masculinity (in our ezdata file, there are actually 110 men and 118 women).

The crosstabulation table of this example would have four cells that comprise a matrix of the four possible combinations of the two levels of the two variables. The data in the cells would be the frequencies (i.e., the number) of men and women who were classified as either low or high in masculinity.

The table would present these four cells as a **2 x 2 contingency matrix**. A contingency matrix classifies individuals into a given cell contingent upon their exhibiting a particular *combination* of one level of the first variable combined with one level of the second variable (e.g., being *both* a man and in the high masculinity category).

If there is *no* relationship between **gender** and **masc** (i.e., if they are independent), then the printout would show just as many low-masculine men as high-masculine men, and the same would hold for female employees. Further, among high-masculinity employees, there would be an equal frequency of men and women, and the same would be true for low-masculinity employees. In other words, the 100 men and women would distribute themselves evenly across the four cells. Thus, the crosstabulation of frequencies would show equal frequencies in all four cells of this contingency table (25 per cell) as shown in Table 10.1.

Table 7.1				
		Gender		
		Male	Female	**Total**
Masculinity	Low-masculine	25	25	50
	High-masculine	25	25	50
	Total	50	50	100

In looking at this hypothetical table, we can see that, indeed, there is no pattern of frequencies beyond what would be expected by chance, indicating that the two variables are independent of each other. Specifically, looking down the *columns* of Table 10.1, we see that among the total of 50 men, 25 are low-masculine and 25 are high-masculine. The same is true for women - 25 are low-masculine and 25 are high-masculine. Thus, from this hypothetical crosstabulation, we would conclude that men are equally-likely to be low-masculine or high-masculine, and so are women.

Further, looking across the *rows*, among the total of 50 low-masculine employees, 25 are men and 25 are women. And for the 50 high-masculine employees, 25 are men and 25 are women. Thus, we would conclude that men are equally likely as women to be low in masculinity, and that women are equally likely as men to be high in masculinity. Again, this is the pattern of frequencies that would be expected by chance alone if the two variables are unrelated, as is the case in thistable.

However, if the hypothesized relationship really exists, then there would be a pattern of frequencies that is different from chance expectations. A hypothetical example of a pattern reflecting a real relationship between **gender** and **masc** is illustrated in Table 10.2.

Table 7.2				
		Gender		

		Male	Female	**Total**
Masculinity	Low-masculine	10	40	50
	High-masculine	40	10	50
	Total	50	50	100

It can be seen from the hypothetical data in Table 10.2 that there is a higher frequency of men who are high-masculine (40 out of 50) than men who are low-masculine (only 10 out of 50). Further, there is a higher frequency of women who are low-masculine (40 out of 50) than women who are high-masculine (only 10 out of 50). It can also be seen that among the 50 high-masculine employees, there is a higher frequency of men (40) than women (10). Last, among the 50 low-masculine employees, there is a higher frequency of women (40) than men (10).

Thus, this table shows that there is a clear relationship between an employee's gender and his/her masculinity level. Masculinity level varies systematically across gender, with men being more likely than women to be high-masculine and women being more likely than men to be low-masculine.

Of course, the above examples were made to be very clear cut. Bivariate frequency distributions often do not lend themselves to such an easy visual determination of whether the frequencies indicate that the two variables are related or not. Only by running the crosstabulation procedure and computing the appropriate statistical test could we answer the question about this hypothesized relationship between gender and masculinity. The statistical test of interest here is called ***Chi square***.

As we will see, SPSS will compute a ***Pearson chi square*** value that will answer the question of whether the actual data in our ezdata file demonstrate a **statistically significant** relationship (i.e., a *real* one) between gender and masculinity, or if they will show that the two variables are statistically independent (i.e., any apparent pattern in frequencies is not real, but due to *random chance*).

Chi square is computed based on a comparison of actual frequencies observed in our sample to that which would be expected to occur by chance alone. If there is a large difference between the observed vs. the expected frequencies, a large value for ***Chi square*** will be obtained. More importantly, the ***probability*** associated with this ***Chi* square** value is computed. This value determines whether the *chi square* value is statistically significant. The general convention used by researchers is that if this probability is **.05** or lower, then we reject random chance as an explanation, and conclude that this is a real (statistically significant) relationship.

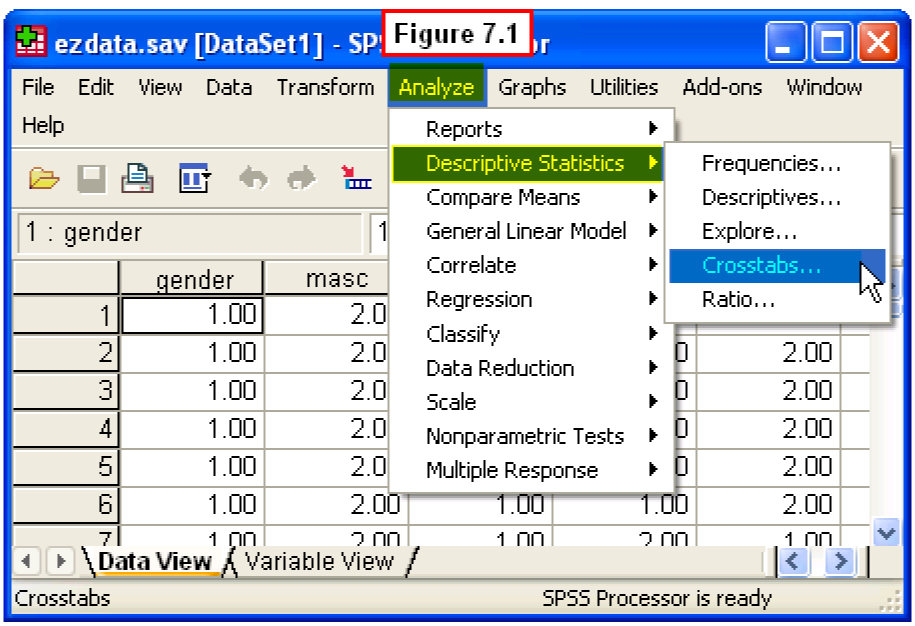
Thus, ***Chi square*** is an inferential statistic - it allows us to make inferences from our sample to the population regarding the hypothesized relationship. This process begins with the assumption that the relationship is due to chance - this is called the **null hypothesis** (**Ho**). Based on the obtained probability, we either retain or reject the null hypothesis using the following decision rule:

- If the probability is < or = .05, reject **Ho** and conclude the relationship is significant
- If the probability is > .05, retain **Ho** and conclude the relationship is due to chance.

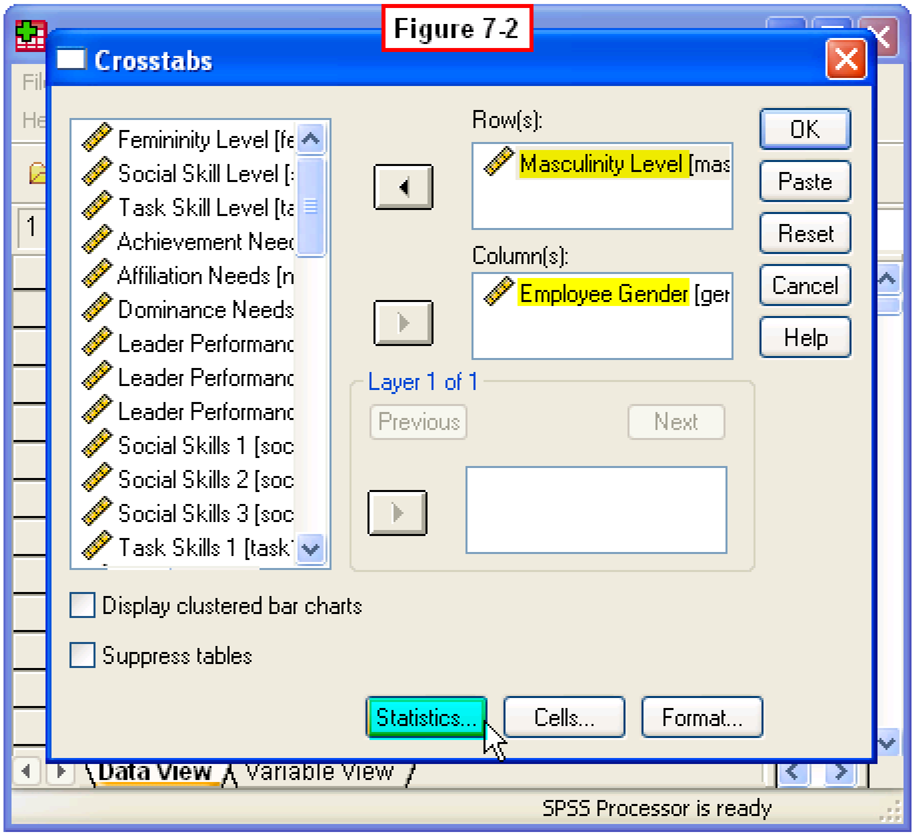
Thus, if the obtained probability is less than or equal to .05, we would conclude that the pattern of frequencies discussed in the crosstabs table is a real one (not due to chance), indicating that masculinity level is, indeed, significantly related to gender.

7.2 Running the Crosstabs Procedure

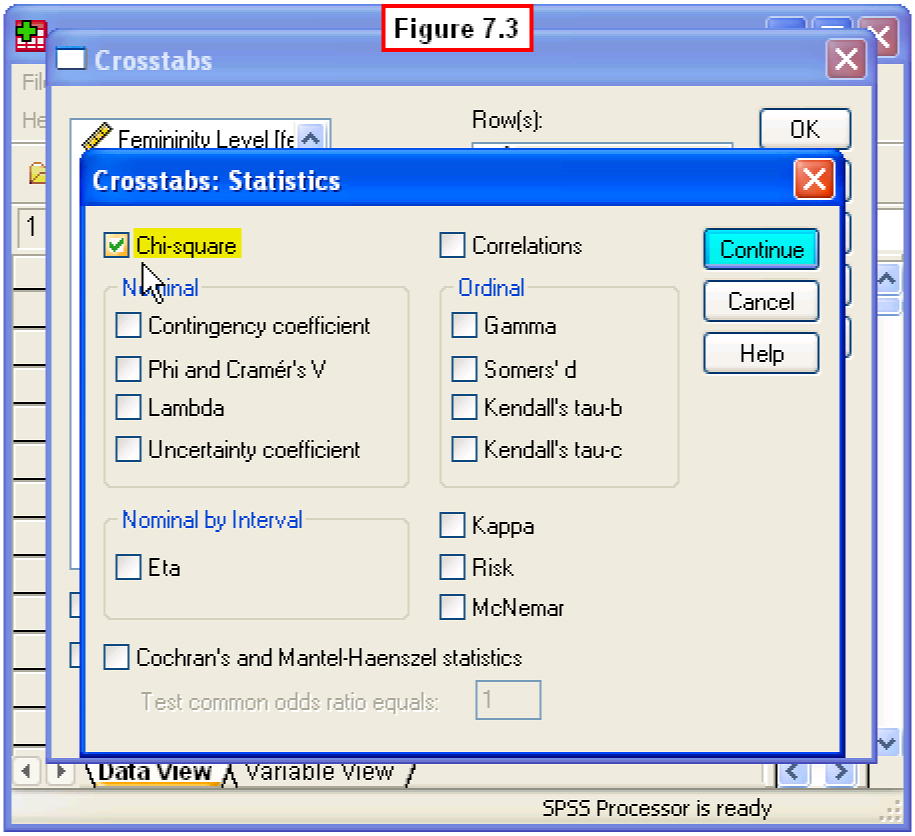
To begin, open **ezdata.sav** and select **Analyze, Descriptive Statistics, Crosstabs...** from the Data Editor menu (Figure 7.2).



Select **Employee Gender** In the left pane of the **Crosstabs** dialog window (Figure 7.2), then click the right-arrow between panes to move this variable into the **Column(s)** pane on the right. Select **Masculinity Level** in the left pane and move this variable to the **Row(s)** pane on the right. Click the **Statistics...** button at the bottom of this window.



Click the **Chi-square** checkbox in the **Crosstabs: Statistics** dialog window (Figure 7.3), then click the **Continue** button on the right.



[Show Me Video!](#)

7.3 Interpreting the Output

The first table of the resulting output file simply lists the number of cases processed (228 total employees). The next table in the output presents the 2 x 2 contingency matrix, or the crosstabulation of frequencies listing the number of male/female employees in the low/high masculinity categories (Figure 7.4). The frequencies highlighted in yellow constitute the four cells of this matrix, and we will need to interpret these if the **Chi square** test reveals that the variables of **gender** and **masculinity level** are significantly related.

For now, looking in the column totals, we see that of the 228 employees 110 are male and 118 female. Looking in the row totals we see that of the 228 employees, 88 are in the low-masculinity category and 140 are in the high-masculinity category. Thus, there is a fairly equal number of male and female employees, and in general, employees are much more likely to be high-masculine than low-masculine.

Figure 7.4

Masculinity Level * Employee Gender Crosstabulation

Count

		Employee Gender		
		Male	Female	Total
Masculinity Level	Low-Masculinity	29	59	88
	High-Masculinity	81	59	140
Total		110	118	228

Examination of the *pattern* of highlighted yellow frequencies in the *cells* of the matrix involves:

- comparing the number of men vs. women in within masculinity levels (the horizontal red double-arrows), then
- comparing the number of low- vs. high-masculine employees within gender (the vertical blue double-arrows).

Recall that if these two variables are independent (i.e., not significantly related), then the frequencies would be relatively evenly distributed across masculinity level and gender. That is, there would be relatively equal numbers of men and women in the low-masculinity category as well as in the high masculinity category. Further, there would be an equal number of low-masculine men as high-masculine men, and an equal number of low-masculine vs. high-masculine women.

Scanning these four frequency counts, it appears that the above is generally not true. Rather, there is a pattern of frequencies across the four categories that indicates unequal frequencies. Before we can interpret this apparent pattern, however, we need to examine the **Chi square** statistic to determine whether or not that this pattern reflects a statistically significant relationship. The **Pearson Chi-square** value computed by SPSS in shown in the last table of the output (Figure 7.5).

Figure 7.5

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13.420[b]	1	.000		
Continuity Correction[a]	12.441	1	.000		
Likelihood Ratio	13.626	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	13.361	1	.000		
N of Valid Cases	228				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 42.46.

The **Chi square** value is 13.42. Recall that to determine whether or not this value indicates a significant relationship, we need to examine the **probability** that this distribution of frequencies occurred by chance alone. Recall that the conventional probability level used to answer this question is **.05** and the following decision rule is employed:

- if the probability is **greater than .05**, then the variables are **not** significantly related
- if the probability is **less than or equal to .05**, then the variables **are** significantly related.

Instead of using the term, **probability**, SPSS uses the term, **significance**, and abbreviates it as **Sig.** More specifically, in Figure 7.5, the **Assym. Sig.** (2-sided) column lists the probability of interest. The probability is **.000** in this column for the **Pearson Chi-square** statistic.

Note that this means the probability is actually something lower than .0005 (SPSS rounds off to three decimal places). Thus, while the actual probability may not be *exactly* equal to zero, it is certainly less than the cut-off value of .05. So since the probability is less than .05, we can reject the null hypothesis of chance as an explanation and conclude that there *is* a statistically significant relationship between **gender** and **masculinity**. This justifies our interpretation of the pattern of frequencies in the cells of Figure 7.4. Examination of this matrix in Figure 7.4 reveals that the pattern is more similar to that shown in Table 7.2 than that in Table 7.1.

Starting with the comparisons between men and women (the horizontal red double-arrows), we see that of the 88 low-masculine employees, 59 are women, while only 29 are men. Thus, there were more female EZ employees than male employees in the low masculine category. Further, of the 140 high-masculine employees, 81 are men compared to 59 women. Thus, there were more male employees than female employees in the high-masculine category.

The above pattern follows social stereotypes about sex role identity. That is, given the emphasis on masculinity in the socialization process, it is perhaps not surprising that our sample would have more low-masculine women than men, and more high-masculine men than women.

However, comparisons within genders (the vertical blue double-arrows) reveal a different and interesting pattern. On the one hand, social stereotypes can still be seen for men: of the 110 men, 81 were in the high-masculinity category, compared to only 29 in the low-masculinity category. Thus, the majority of male employees were high-masculine.

On the other hand, the comparison within female employees disconfirms social stereotypes: of the 118 women, half were low-masculine (59) and half were high-masculine (59). Thus, at least at EZ Manufacturing, there were just as many high-masculine as low-masculine women. This illustrates the truism that research and data analyses sometimes confirm hypotheses and expectations, but sometimes they yield surprises - this one reason research and data analysis can be so interesting!

There are a variety of possible explanations of the large number of high-masculine female employees, some of which we will discuss in later chapters. For now, one possible explanation is that EZ is a high-tech manufacturing firm, and technological-orientation is stereotypically associated with masculinity in our culture. Thus, perhaps female EZ employees either adapted their personalities to fit the masculinity stereotype, or perhaps high-masculine women self-selected to work at EZ rather than in more traditionally feminine work roles.

We will gain more insights into the relationships among variables in from our project in Chapter 8, where we introduce a commonly used analytical approach known as correlation. This type of analysis also concerns bivariate distributions, but rather than examining frequencies to assess relationships (as in **Crosstabs**), correlation directly assesses covariation in scores on two variables.

7.4 Chapter Review Video

Review Me!

7.5 Try It! Exercises

1. Examining the relationship between Gender & Femininity Level using Crosstabs

In the chapter example we saw there was a relationship between gender and masculinity level among EZ employees. Is there a relationship between gender and femininity level? Use the **Crosstabs** procedure described in Section 7.2 to create a crosstabulation of frequencies on the variables, **Gender** and **Femininity Level**.

- In the **Crosstabs** dialog window, move **Employee Gender** to the **Column(s)** box, and move **Femininity Level** to the **Row(s)** box. Then click the **Statistics** button.
- Check the **Pearson Chi-square** box in the **Crosstabs: Statistics** dialog window, then click the **Continue** button.
- Click the **OK** button in the **Crosstabs** window. The crosstabulation and **Chi square** tables will appear in an Output Viewer window.
- **Print** your output file to submit to your instructor.
- **Write an interpretation of the results** to submit to your instructor. Follow the example in Section 7.3, stating your conclusions:
    - Interpret the **Pearson Chi-square** value; is **Gender** significantly related to **Femininity Level**, or are these variables independent of each other?
    - State the relative number of male vs. female employees in the low-femininity category
    - State the relative number of male vs. female employees in the high-femininity category
    - State the number of males who are in the low- vs. the high-masculinity category
    - State the number of females who are in the low- vs. the high-masculinity category
- **Write a statement explaining** these patterns of frequencies. That is, how might these patterns be explained in terms of social stereotypes about gender and femininity level?

Chapter 8
Correlation:
Understanding Bivariate Relationships Between Continuous Variables

8.1 Introduction to the Pearson Correlation Coefficient: *r*

In Chapter 7 we demonstrated how to use the **Crosstabs** procedure to examine the relationship between pairs of categorical variables. As part of this procedure, we also discussed how we could use the statistical measure of association, ***Chi square**,* to determine whether or not a relationship between two categorical variables is statistically significant.

Recall from Chapter 5 that we created these categorical variables by transforming variables that were originally continuous. For example, scores on the continuous variable, **masctot**, range from 5 to 35. These scores were transformed into a categorical variable, **masc**, with only two levels (1= low masculinity; 2 = high masculinity). While such transformations permitted the construction of contingency tables for crosstabulation, there are some disadvantages to converting continuous data to categorical data.

Such transformations result in a loss of detail and precision that might affect measures of association between variables. Statistically, the reduction in variability in the transformed categorical variables can decrease our ability to assess relationships. There are other ways to evaluate the relationship between continuous variables, and one such procedure involves the calculation of the **Pearson correlation coefficient**, known as ***r***.

In general, the ***Pearson correlation coefficient*** is a statistic used to determine the ***degree*** and ***direction*** of relatedness between two continuous variables. The possible values of the correlation coefficient range from -1.00 to +1.00, and the closer the number is to an absolute value of 1.00, the greater the degree of relatedness. As with ***Chi square***, the ***Pearson correlation coeffficient*** can be tested for statistical significance (using the conventional probability criterion of **.05**).

Two variables may be strongly related, as is the case for age and body weight before puberty (10 year olds invariably weigh more than 5 year olds). Or the two variables may be weakly related, as is the case for these same two variables, age and body weight, between ages 30 and 50 (45 year olds are not likely to weigh more than 35 year olds).

Sometimes changes in one variable are associated with no predictable change in the second variable. For example, in adults height and intelligence are unrelated: tall people can be either smart or dull, and the same holds for short people. In such cases the correlation between the variables centers around zero; that is, there is no correlation.

The *direction* of the relationship, either positive or negative, is indicated by the sign (+ or -) of the correlation value (i.e., whether the coefficient is a positive or negative number). High school SAT scores and college GPA's are positively correlated: high values on one variable (SAT score) are associated with high values on the second (college GPA). When two variables are negatively correlated, an increase in the first is associated with a decrease in the second. For example, we typically find that the higher one's alcohol intake, the lower that person's motor coordination becomes.

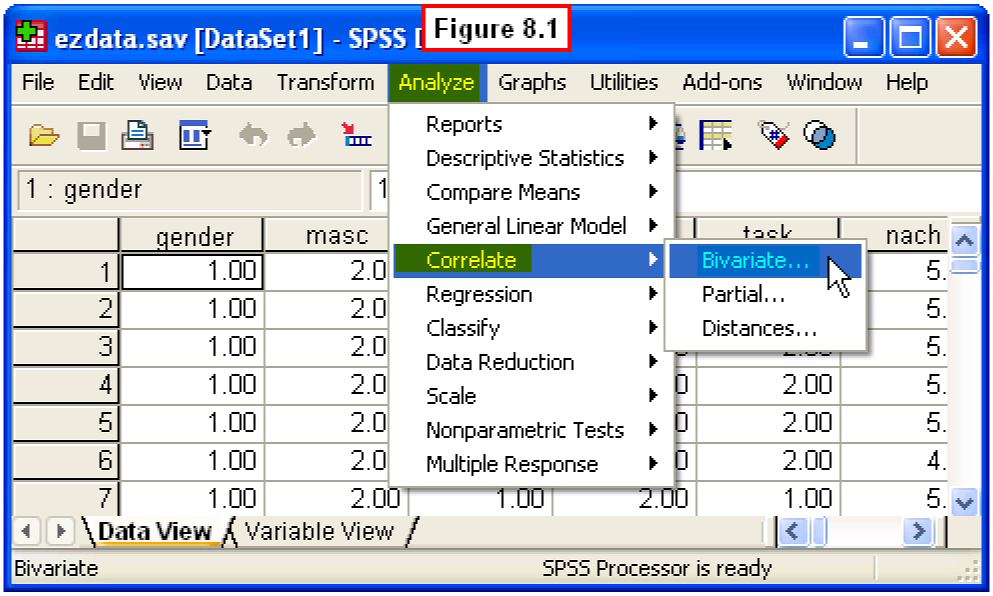
The Pearson correlation coefficient (***r***) is used specifically to describe relationships when the variables to be correlated are continuous (measured on at least an interval scale). This procedure also assumes that the correlated variables are normally distributed, and that the relationship between the two variables approximates a linear one. Curvilinear relationships are best described with other correlational procedures, most of which are beyond the scope of this book.

8.2 Running Correlations

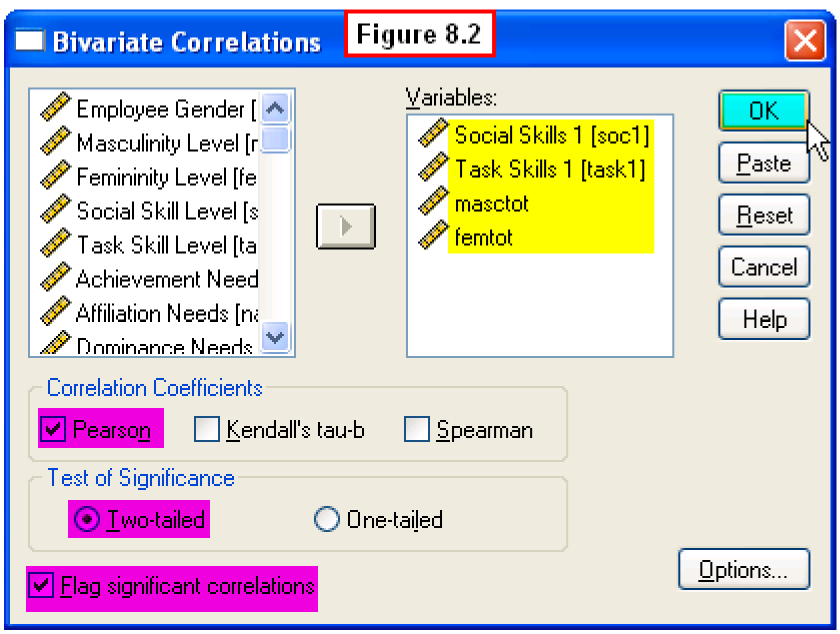
There are numerous bivariate relationships that could be explored using the correlation procedure. For this example we will examine relationships between **masctot** and **femtot** and two other variables, **task1** and **soc1**. In the exercise at the end of the chapter you will be asked to run correlations between **masctot** and **femtot** and two other variables, **naff** and **ndom**. Before we run the analyses, take a moment to think about what kinds of relationships we might expect to obtain here.

For example, given social stereotypes, we might expect a significant positive correlation between degree of masculinity and degree of task skills (i.e., greater masculinity might be related to higher degree of task skills). For the same reasons we might expect a significant positive correlation between degree of femininity and social skills. However, we might find that masculinity is either unrelated or negatively correlated with social skills, and that femininity might be unrelated or negatively correlated with task skills. Actually, the procedure we will use will generate all possible intercorrelations among the included variables.

To run the correlations, open **ezdata.sav** and select **Analyze, Correlate, Bivariate...** from the Data Editor window (Figure 8.1)



In the **Bivariate Correlations** dialog window (Figure 8.2), select **Social Skills 1(soc1)**, **Task Skills1 (task1)**, **masctot** and **femtot** from the variables in the left pane, then click the right-arrow key to move these four variables to the **Variables:** pane on the right. Note that you will have to scroll down the left pane to locate these variables.



Note that by default **Pearson** is checked for the desired correlation, **Two-tailed** is selected for the test of significance, and the **Flag significant correlations** option is checked. Leave these settings as they are. Then click the **OK** button in the upper right corner. The requested correlations will appear in an Output Viewer window.

[Show Me Video!]

8.3 Interpreting the Output

The output file contains a matrix of all possible correlations among the variables (Figure 8.3).

Figure 8.3

Correlations

		Social Skills 1	Task Skills 1	masctot	femtot
Social Skills 1	Pearson Correlation	1	.301**	.124	.622**
	Sig. (2-tailed)		.000	.061	.000
	N	228	228	228	228
Task Skills 1	Pearson Correlation	.301**	1	.488**	.283**
	Sig. (2-tailed)	.000		.000	.000
	N	228	228	228	228
masctot	Pearson Correlation	.124	.488**	1	-.063
	Sig. (2-tailed)	.061	.000		.346
	N	228	228	228	228
femtot	Pearson Correlation	.622**	.283**	-.063	1
	Sig. (2-tailed)	.000	.000	.346	
	N	228	228	228	228

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

The four variables (highlighted blue) are listed in rows as well as in the columns (thereby creating the *matrix* of all possible correlations). In each cell (or "box") of this matrix are three rows.

- The first row contains the **Pearson *r*** value (highlighted yellow)
- The second row contains the two-tailed probability (**Sig.**, highlighted purple)
- The third row shows the number of pairs of scores on which each r value is computed.

Next note that the ***r*** value for the correlation between each variable and itself (e.g., **Social Skills 1** with **Social Skills 1**) equals 1. This is because the correlation of any variable with itself is perfect - something of no interest, so these boxes can be ignored.

Also note that the 1's form a diagonal cutting from the upper left to the lower right of this matrix. *Note also that all of the correlations listed below the diagonal are duplicates of the correlations shown above the diagonal*. So we need only pay attention to the six correlations shown above the diagonal to interpret our output, since the information below the diagonal is redundant. Last, note that asterisks are placed next to the ***r*** values for which the probability is less than or equal to .01 to flag these as statistically significant correlations.

Now we can begin interpreting the **degree** and **direction** of the relationships indicated by the six correlation coefficients of interest. Recall that the **direction** is easily seen by the sign of the ***r*** value (+ or -). A positive correlation indicates that high scores on one variable are associated with high scores on the second variable, while a negative correlations indicates that high scores on one variable are related to low scores on the second variable. All of our significant correlations are positive in this example, so we will discuss the more complex issue of interpreting degree of relationship.

There are two ways to interpret the degree of relationship. The first is an all or none judgment of whether or not the variables are significantly related. The null hypothesis (**Ho**) is that ***r*** = **0.0**, meaning that the variables are unrelated (hence, the **degree** of relationship is zero). Recall the convention is:

- If the **Sig.**, or probability (*p*), associated with the ***r*** value **.05** or less, then we reject **Ho**, and conclude that there is a statistically significant relationship between pair of variables.
- If $p > .05$, then we retain **Ho**, and conclude that the variables are unrelated.

Thus, if the **Sig.** value listed for a correlation is .05 or less, we can assume that the correlation is not the result of chance or random sampling error. That is why we would reject **Ho** and conclude that the correlation is a real one, and thus, one that can be generalized from the sample to the overall population in which we are interested.

So to interpret our correlation matrix, we first need to look at the **Sig.** values (*p*) listed below each ***r*** value (recall that SPSS makes this easier by flagging correlations for which $p < .05$, placing asterisks next to the ***r*** values). Examining these, we see that four of the six correlations are significant. The nonsignificant correlations are between **Social Skills** and **masctot** ($r = .12$, $p = .06$), and between **masctot** and **femtot** ($r = -.06$, $p = .35$). Thus, our first conclusions concern the nonsignificant correlations. We can conclude that employees' social skills are unrelated to their masculinity, and their degree of femininity is unrelated to their masculinity.

The latter conclusion is of theoretical interest, because intuitively one might have expected masculinity and femininity to be significantly correlated in a *negative* direction (high masculinity would be related to low femininity). The fact that the correlation is not significant supports the proposition that these variables are independent, and that it is possible to have all combinations of low/high levels of masculinity and femininity.

Recall that the concept of *androgyny* implies that these variables are independent. That is, it possible to score high or low on one variable and either high or low on the other. For example, androgynous individuals score high on both, while sex-typed persons score high on one, but low on the other). Thus, this nonsignficant correlation supports the theory of androgyny.

Turning to the four significant correlations, note that although they are all significant, the ***r*** values vary widely - the coefficients range from .68 (the strongest) to .28 (the weakest). Beyond the all-or-none decision about statistical significance, researchers also often elaborate on the degree of relationship. While there are no hard-fast rules, here is a general rule of thumb:

- ***r*** values greater than **.50** indicate a **strong** correlation
- ***r*** values around **.30** indicates **moderate** correlation
- ***r*** values less than **.20** indicate a **weak** correlation

The import of this is that it is possible for two variables to be significantly related (statistically), even if the relationship is a weak one. Thus, researchers typically expect a statement of degree to modify the basic conclusion that a relationship is statistically significant.

Using the above guidelines looking across the first row of the matrix, we can conclude that there is a moderate positive relationship between social skills and task skills (r = .30, p = .000), and a strong positive correlation between social skills and femininity (r = .62, p = .000). Thus, a higher degree of social skills is related to both a higher degree of task skills and higher femininity, but the latter is a much stronger relationship.

Looking in the second row, we can conclude that there is a moderate positive relationship between task skills and femininity (r = .28, p = .000), and a strong positive correlation between task skills and masculinity (r = .49, p = .000). Thus, a higher degree of task skills is related to both higher femininity and higher masculinity, but the latter is a much stronger relationship.

So these analyses have confirmed some expectations, but not others. It is perhaps not surprising that the strongest positive correlations were obtained between masculinity and task skills and between femininity and social skills. Task skills in leadership are associated with instrumentality, a stereotypically masculine characteristic. And Social skills in leadership are associated with expressiveness, a stereotypically feminine characteristic.

However, it is interesting to note that although masculinity was unrelated to social skills, the correlations between femininity and task skills was moderate and statistically significant. Thus, masculinity was strongly related to task skills, and masculinity was independent of social skills (as one might have intuited). However, while femininity was strongly related to social skills, it was not unrelated to task skills (as one might have expected based on social stereotypes).

Confused? We are still in the preliminary stages of identifying the pieces of the puzzle, so we don't yet have the larger picture in front of us. The relationships will become clearer as we carry out other kinds of analyses in the succeeding chapters. However, this first run has generated some interesting correlations that suggest what kinds of characteristics seem to go together in EZ employees. You will discover some similar types of relationships in the exercise to follow, and you are encouraged to explore other correlations in the data file on your own!

8.4 Chapter Review Video

Review Me!

8.5 Try It! Exercises

1. Correlations among Affiliation needs, Dominance needs, Masculinity & Femininity

In the chapter example we examined relationships between masculinity/femininity and task/social skills. Are there also relationships between masculinity/femininity and need for dominance and need for affiliation? Use the **Pearson correlation** procedure described in Section 8.2 to generate a correlation matrix of the variables, **Affiliation Needs (naff)**, **Dominance Needs (ndom)**, **masctot** and **femtot**.

- In the **Bivariate Correlations** dialog window, move the above four variables from the left pane to the **Variables:** pane on the right.
- Make sure Note that the default **Pearson** is checked for the desired correlation, **Two-tailed** is selected for the test of significance, and the **Flag significant correlations** option is checked.
- Click the **OK** button. The requested correlation matrix will appear in an Output Viewer window.
- **Print** your output file to submit to your instructor.
- **Write an interpretation of the results** to submit to your instructor. Follow the example in Section 8.3, stating your conclusions:
  - Interpret the **Pearson r values**; remember to state which variables are significantly related and which ones are unrelated.
  - State the **degree** and **direction** of relationship among those correlations that are significant (*p* less than or equal to .05).
- **Write a statement explaining** these patterns of relationships. That is, how might these patterns be explained in terms of social stereotypes about femininity, masculinity, need for affiliation and need for dominance?

Chapter 9

Independent Samples t-Test: Assessing Differences Between Two Independent Group Means

Jennifer Winquist
Valparaiso University

9.1 Introduction to the t-Test

In the previous chapters we examined *relationships* among variables to assess covariation between the variables. In this chapter we consider research concerning *differences* between groups. Experiments are designed to establish cause-effect relationships. In the simplest experiment there are two groups of participants created by the manipulation of an independent variable (the cause). The two groups are measured on the same dependent variable (the effect) in order to compare their scores. Data analyses are needed to determine whether the independent variable manipulation produced significant differences in scores between the two groups on the dependent variable.

The t-test is used to determine whether the difference between means of two groups or conditions is due to the independent variable, or if the difference is simply due to chance. Thus, this procedure establishes the probability of the outcome of an experiment, and in doing so enables the researcher to reject or retain the null hypothesis (in this case, **Ho** is that any observed differences are *not* significant, but rather, are due to chance).

The null hypothesis states that the experimental manipulation has no effect, therefore the means of the groups will be equal. In this respect, the t-test is an inferential statistic used to test hypotheses. Under ideal conditions, these types of inferential statistics allow the researcher to infer a causal relationship between the independent and dependent variable.

There are two distinct applications of the t-test. When a **between-subjects design** is used, the **independent-samples t-test** is the appropriate test. Use of a **within-subjects design** (sometimes called a *repeated measures* design) or a participant-by-participant matched design requires analysis with the paired samples t-test (also known as the *correlated* or **paired-samples t-test**). In this chapter we will introduce the independent samples t-test. We will address the correlated samples t-test in the next chapter.

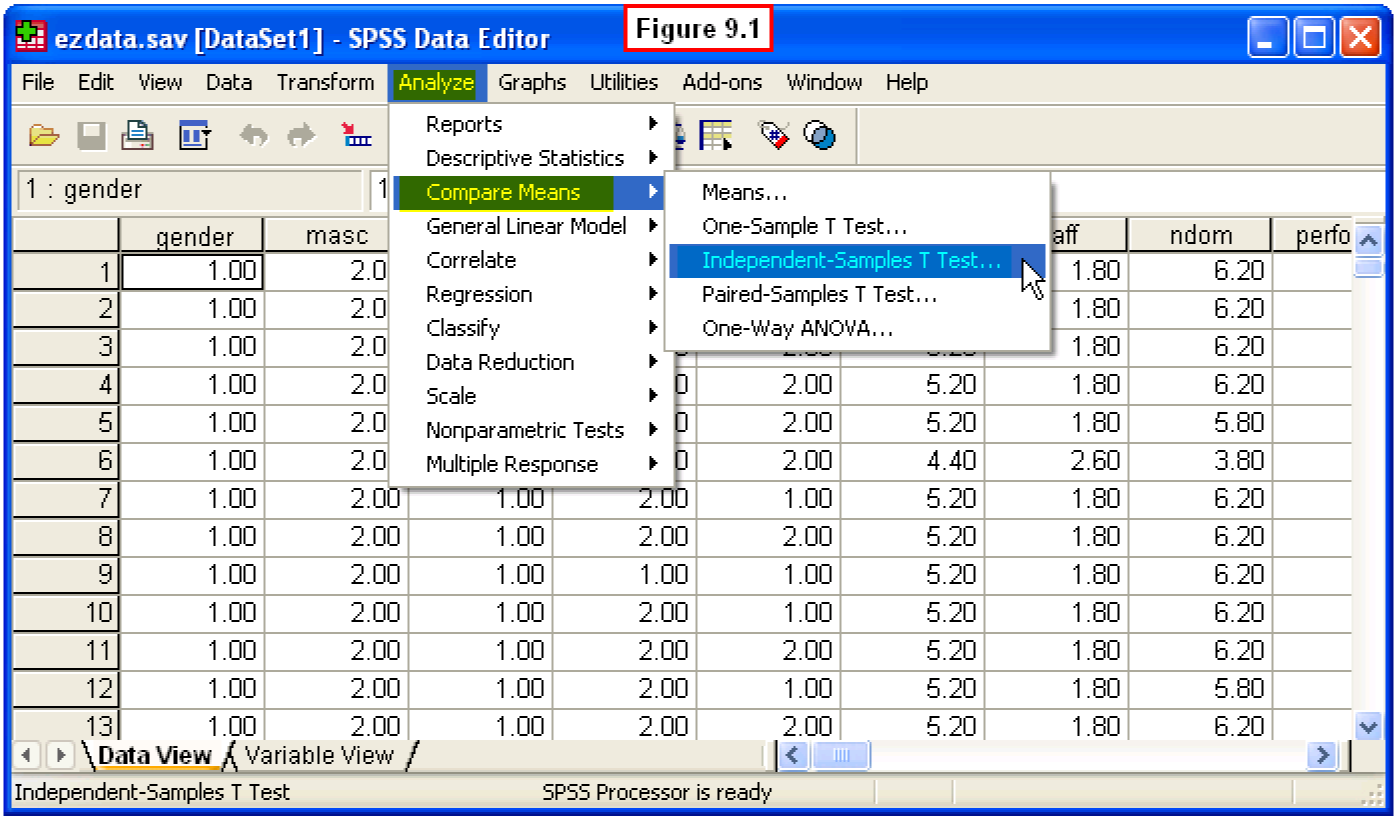
There are many hypotheses we could test as part of this project using the t-test. For example, we are interested in any **gender** differences that might exist among EZ employees. Note that in this context, gender is considered to be a *quasi-independent* *variable* because we cannot actually manipulate gender. Nevertheless, the **t-test** can be applied to examine differences between men and women on various dependent variables. For example, we might want to test for gender differences in scores on the task skills scale and/or the social skills scale (both would be considered dependent variables in this context).

You might hypothesize that men would score higher than women on task skills, because research indicates that task orientation in leadership is a stereotyped male characteristic. Comparing the scores of women on task skills against those of men on independent-samples t-test would enable you to determine whether this was indeed the case. On the other hand, you might expect women to exhibit greater social skills (a stereotyped female characteristic) than men, a hypothesis that could also be tested with the independent-samples t-test. But let's not stop there.

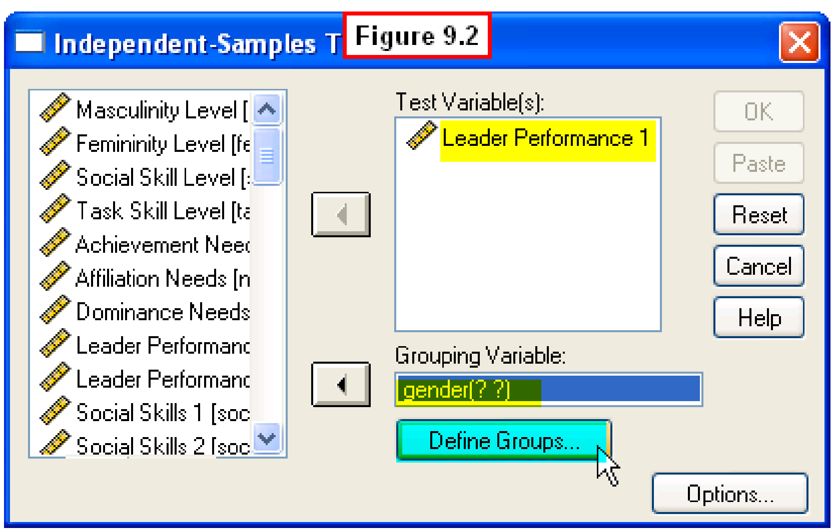
Ultimately we're interested in understanding leadership performance (**perform**) of employees, so why not determine if there are **gender** differences in overall performance? Although in this instance there may be no clear *a priori* reason to suspect that there should be gender differences in performance, it is certainly a question worth investigating. We will use the independent samples t-test to examine gender differences in performance for the example in this chapter. You will be asked to examine gender differences in social skills for the exercise at the end.

9.2 Running the t-test

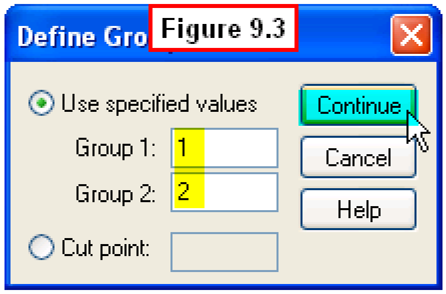
Open your **ezdata.sav** file in the Data Editor. Select **Analyze, Compare Means, Independent-Samples T Test...** from the menu (Figure 9.1).



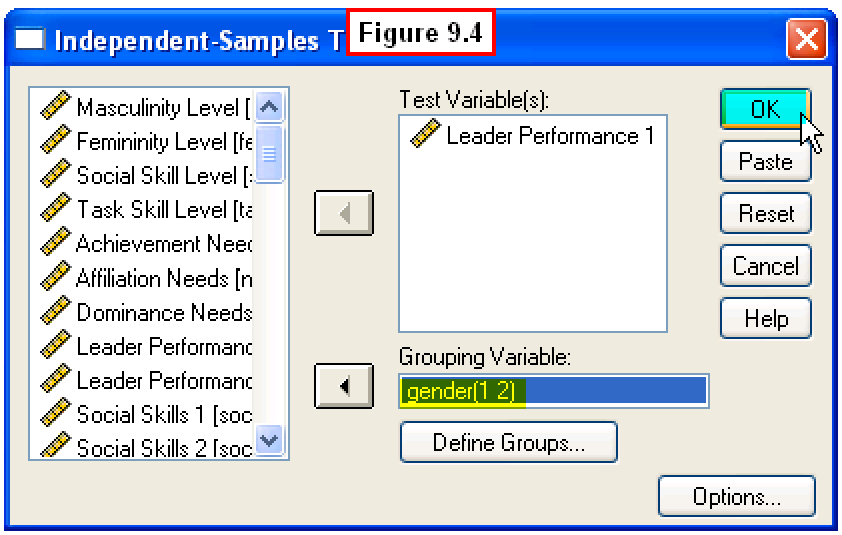
The **Independent-Samples T Test** dialog window will appear (Figure 9.2). In the left pane select our dependent variable, **Leader Performance 1 (perform1)**, and move it to the **Test Variable(s):** pane on the right by clicking the arrow button between the panes. To choose the independent variable, **Employee Gender (gender)**, select it in the left pane and move it to the **Grouping Variable:** pane.



Now, we need to define the two groups to be compared. To do this, click the **Define Groups...** button below the dialog window. The **Define Groups** dialog window will appear (Figure 9.3).



The two groups we want to compare are males and females. Females were given a code of 1, so type **1**in the **Group 1:** box. Males were given a code of 2, so type **2** in the **Group 2:** dialogue box. Next, click the **Continue** button. This will take you back to the **Independent Samples T Test** window. To run the t-test simply click on the **OK** button (Figure 9.4).



Show Me Video!

9.3 Interpreting the Output

The first table, **Group Statistics**, is shown in Figure 9.5. This table includes descriptive statistics for each group. Specifically, the table includes the number of cases (N), the mean leader performance score, the standard deviation, and the estimated standard error of the mean (the standard deviation divided by N).

Figure 9.5

Group Statistics

	Employee Gender	N	Mean	Std. Deviation	Std. Error Mean
Leader Performance 1	Male	110	5.6818	2.74257	.26149
	Female	118	6.1441	1.94046	.17863

Of greatest interest here are the mean performance scores for men (5.68) and for women (6.14). You might be tempted to conclude that this indicates that women had significantly higher average performance scores than men. However, this would be premature - in fact, the whole point of the t-test is to determine whether this is a real difference (statistically significant), or one that could be attributed to random chance. To do this, we need to examine the next table, **Independent Samples Test** (Figure 9.6).

Figure 9.6

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Me			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Leader Performance 1	Equal variances assumed	42.802	.000	-1.477	226	.141	-.46225
	Equal variances not assumed			-1.460	194.923	.146	-.46225

9.3a Testing for Homogeneity of Variance

The first two columns labeled **Levene's Test for Equality of Variances** provides a test of one of the assumptions of the t-test, i.e., that the variance in the two groups are equal (i.e., similar or homogeneous). If this assumption is violated in the data, a statistical adjustment needs to be made. The ***F*** statistic in the first column and its probability in the second column (**Sig.**, an abbreviation for significance) provides this test. If the probability of the ***F*** value (i.e., **Sig.**) is less than or equal to **.05**, then the variances in the groups being compared are different, and the condition of homogeneity of variance has not been satisfied.

The results of the ***F*** test determine whether to use the **Equal variances assumed** rows or the **Equal variances not assumed** rows in evaluating the t statistic. The decision rule for determining which rows to use is as follows:

- **If the variances for the two groups are equal** (i.e., **Sig. > .05**), then use the output in the ***Equal variances assumed*** rows. These rows represent the more conventional method of evaluating the **t** value based upon degrees of freedom (**df**) equal to the total number of scores minus 2 (this is the method that is described in most introductory statistics or research methods textbooks).

- **If the variances for the two groups are significantly different** (i.e., **Sig. < .05**), then use the output in the ***Equal variances not assumed*** row. Evaluation of the **t** statistic in this row is based upon an adjusted degrees of freedom which takes into account the dissimilar variances in the two groups.

Since the probability (**Sig.** = .000) for the ***F*** value is less than **.05**. Thus, the variances of the two groups are **not equal**, and therefore the output in the **Equal variances *not* assumed** row should be used.

9.3b Testing the null hypothesis: Interpreting the significance of the t-value

To determine if the difference in performance between men and women is significant, we need to look in the columns labeled **t-test for equality of Means**. We are currently only interested in the obtained t-value and its probability, which can be seen in the columns labeled ***t***en **Sig. (2-tailed)**. Looking in the **Equal variances *not* assumed** row, we see a ***t*** value of 1.46. The probability in the **Sig. (2-tailed)**column in the (p = .146) is greater than .05, meaning that we need to retain the null hypothesis of no differences, concluding that there was no significant difference in leadership performance between male and female EZ employees.

The following sentence illustrates how these results would be written according to APA format.

The results indicate that there was no significant difference in performance between women and men, t (195) = 1.46, p = .15. That is, the average performance score of women (M = 6.14, SD = 1.94) was not significantly different from that of men (M = 5.69, SD = 2.74).

Note that while researchers generally are interested in finding "significant differences," sometimes the *absence* of a significant difference is of either theoretical or practical value. That certainly is the case here. In particular, these results indicate that *there is no reliable difference in performance* between men and women at EZ Manufacturing. This is an important because this information may be useful in calming the anxieties of upper-level executives who might adhere to the stereotype that women are less capable of men in leadership situations

9.3 c Additional Information in the t-table

There is additional information in the t-table that might be of use to you. The first is the **Mean Difference**. This is simply the difference between the two means. The **Standard Error of the Mean Difference** is the denominator used in computing the t-statistic. Finally, the**95% Confidence Interval for the Difference** consists of two numbers indicating the lower and the upper bound of the confidence interval. We can be 95% confident that the difference between the two means falls between the lower and upper bounds.

As mentioned, there are numerous other hypotheses we could test using the independent samples t-test on the data from our EZ Manufacturing study. The exercise at the end of the chapter illustrates one of these, and you are encouraged to explore others on your own. In the next chapter we will discuss a similar approach to hypothesis testing using the **correlated samples t-test**.

9.4 Chapter Review Video

Review Me!

9.5 Try It! Exercises

1. Using the t-test to Examine Gender Differences in Need for Affiliation

Use the procedures described in Section 9.2 to determine if male and female employees differ in their level of affiliation motivation. Thus, select **Affiliation Needs (naff)** as the **Test Variable** (the dependent variable), and select **Employee Gender (gender)** as the **Grouping Variable** (the independent variable).

- **Print** your output file to submit to your instructor.
- Answer the following questions for the affiliation variable:
  - What is the mean affiliation score and standard deviation for males?
  - What is the mean affiliation score and standard deviation for females?
  - Does this data violate the assumption of homogeneity of variance? Why?
  - How many degrees of freedom are associated with this test?
  - What is the value of the t-statistic?
  - What is the significance level?
  - Should you reject or fail to reject the null hypothesis of no difference in affiliation needs?
- **Write an interpretation of the results** along with your answers to the above questions to submit to your instructor. Follow the example in Section 9.3.

Chapter 10

Paired Samples t-Test:
Assessing Differences Between Correlated Group Means

10.1 Introduction to the Paired-Samples t-Test

In the previous chapter we distinguished between two distinct applications of the t-test: the independent samples t-test and the correlated samples t-test. Recall that when a **between-subjects design** is used, the **independent-samples t-test** is the appropriate test. This is the procedure introduced in Chapter 9.

Use of a **within-subjects design** (sometimes called a *repeated measures* design) or a participant-by-participant matched design requires analysis with the paired samples t-test (also known as the *correlated* or **paired-samples t-test**). We turn our attention to this application of the t-test in this chapter.

The differences in these types of designs is both procedural and statistical. In the independent samples design, separate (and independent) groups of participants are compared, and each participant is measured only once on the dependent variable. The two sets of scores are therefore independent (uncorrelated).

In the correlated samples design, there are still two sets of scores on the dependent variable, but the scores are not independent. There are three ways in which a design can result in correlated scores:

- **Repeated Measures** are obtained on one group of participants, such as in measuring participants **before** a treatment is applied and again **after** the treatment. Thus, each person serves as his/her own control, and because the two sets of scores to be compared are obtained from the same people, the two groups of scores are not independent.
- Groups are established by **pairing** participants on the basis of some **natural relationship** between the individuals in each pair. For example, EZ employees' social skills might be compared to those of their fathers. Each person would be paired with his/her father in this design, and because of this natural relationship, the scores are not independent.
- A **matched-participant** design is employed in which the participants are paired based on receiving similar scores on some pretest. One person in each pair is then assigned to the experimental group (e.g., the group receiving some treatment) and one in each pair is assigned to the control group (which doesn't receive the treatment). This is sometimes done in order to control for some individual difference variable. For example, we might want to control for intelligence in assessing the effects of a drug on memory. We would first

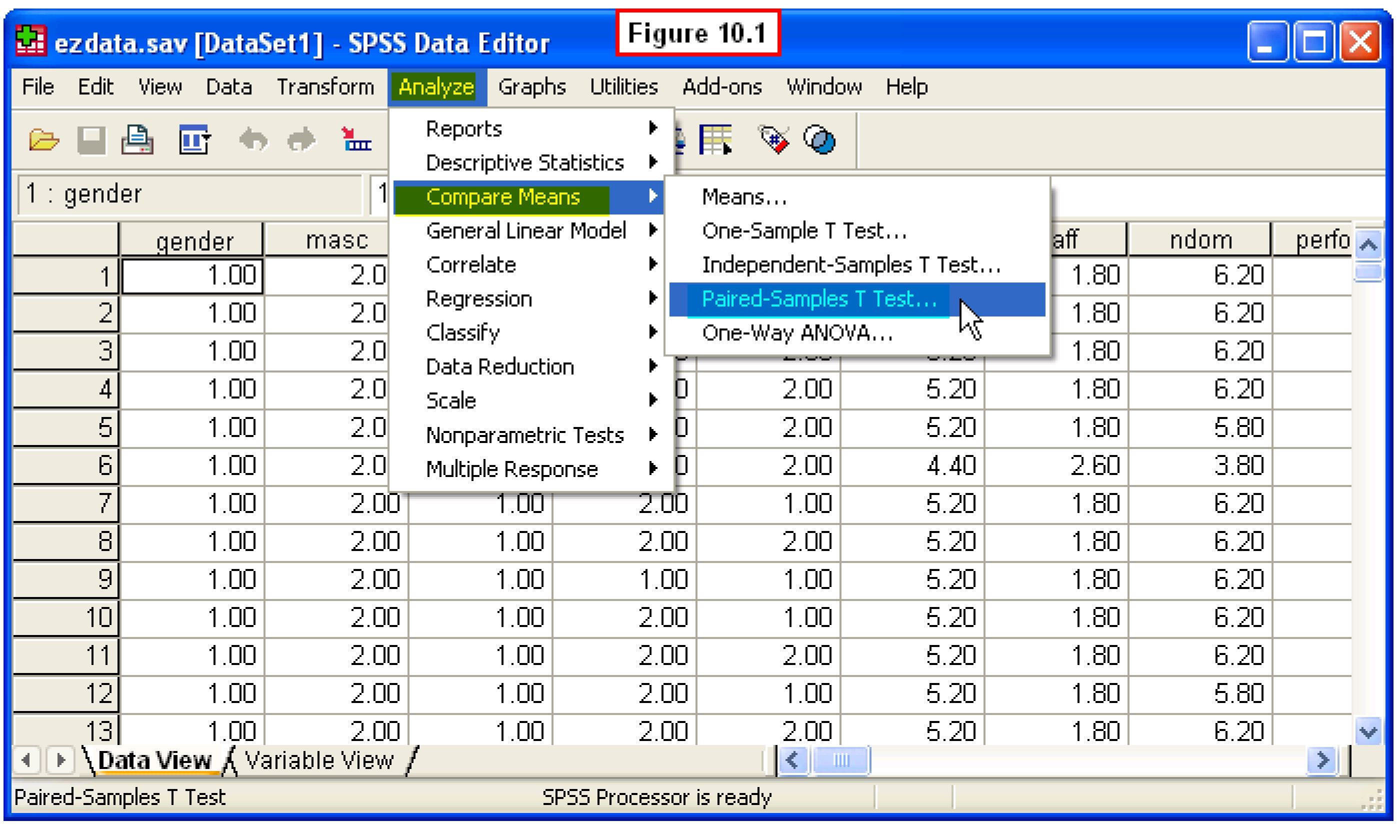
measure the participants' IQ and then match them based on their IQ scores before randomly assigning one to the drug and the other to the no-drug condition.

If any of the above three situations exists, the scores in the two groups will be correlated based on the pairing. This intercorrelation must be accounted for statistically when comparing the two groups, and that is what the **paired-samples t-test** does.

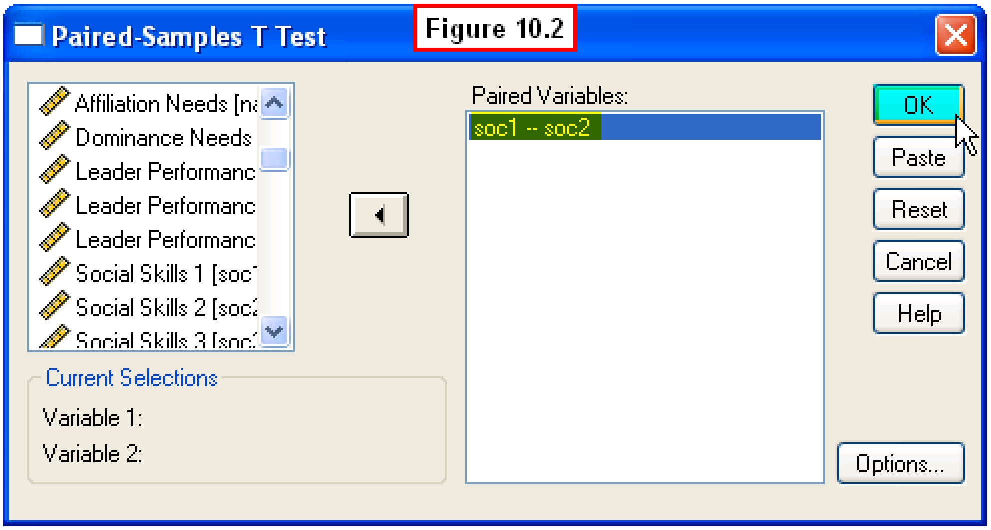
Recall from the EZ Manufacturing project that we do have three **repeated-measures** variables. Participants' **social skills**, **task skills** and **performance** scores were obtained prior to attending a leadership training workshop, then again after the workshop. The **paired-samples t-test** is the appropriate statistical test to determine whether or not there was a significant change in scores on these variables after the workshop compared to before participation. We will use **social skills** for the example in this chapter.

10.2 Running the t-test

Open your **ezdata.sav** file in the Data Editor. Select **Analyze, Compare Means, Paired-Samples T Test...** from the menu (Figure 10.1).



In the left pane of the **Paired-Samples T test** dialog window, select **Social Skills 1 (soc1)** and **Social Skills 2 (soc2)** and move them to the **Paired Variables:** pane on the right by clicking the arrow button in the middle. Then click the **OK** button.



[Show Me Video!]

10.3 Interpreting the Output

Several statistics are presented in the first table, **Paired Samples Statistics** (Figure 10.3). The most relevant for our purposes are the two means. Examination of these means suggests that the average leadership performance score was higher after the workshop. However, the t-test will determine whether or not this difference is real or due to chance.

Figure 10.3

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Social Skills 1	5.4605	228	2.02043	.13381
	Social Skills 2	5.8553	228	2.05242	.13593

The paired-samples t-test procedure automatically computes the correlation between the two sets of scores (Figure 10.4). As discussed in the previous section, we can see that there is, indeed, a significant positive correlation between the scores before and after the workshop (r = .29, p < .001), indicating that those who scored high on social skills before the workshop also tended to score high on social skills after the workshop.

Figure 10.4

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Social Skills 1 & Social Skills 2	228	.288	.000

The next table presents the results of the t-test (Figure 10.5). The first column shows the actual difference between the two means (-.40), which is the numerator of the t-test formula. This is a negative number because the mean after the workshop (a larger value) is subtracted from that before the workshop (a smaller value).

The second column presents the standard deviation of the difference scores (2.43). The third column shows the standard error of the mean (.16), which is the denominator of the t-test formula. The 95% Confidence Interval presents the lower and upper limits of the mean difference (i.e., we can be 95% confident that the difference between the means falls somewhere between -.71 and -.08)

Figure 10.5

Paired Samples Test

		Paired Differences					t	df	Sig.
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Social Skills 1 - Social Skills 2	-.39474	2.43006	.16093	-.71185	-.07762	-2.453	227	

The most important columns for our purposes are the one presenting the obtained t-value (-2.45) and its probability, or significance (.02). Since the probability is less than .05, we reject the null hypothesis of no difference in favor of the alternative hypothesis that the difference is real. To report the results in APA format, we would state the following:

A paired-samples t-test revealed a significant differences in the social skills scores before and after the workshop, t (227) = -2.45, p < .02. This indicates that the mean social skills score after the workshop (M= 5.86) was significantly higher than the mean before the workshop (M = 5.46).

This analysis provides evidence that the leadership training workshop was, indeed, effective in leading to increased social skills in leadership situations. This type of research is essential to evaluate the effectiveness of any new program or procedure implemented, and as such, it is frequently called program evaluation research. In the exercise at the end of the chapter you will be asked to determine whether or not **task**

skills also changed significantly after the workshop. You might also be curious to determine whether or not leader **performance** changed as well.

10.4 Chapter Review Video

[Review Me!]

10.5 Try It! Exercises

1. Using Paired-Sample t-test: Effects of Leadership Training Workshop on Task Skills

Use the procedures described in Section 10.2 to determine if there was a significant change in task skills scores before and after the leader training workshop. Select **Task Skills 1 (task1)** and **Task Skills 2 (task2)** as the **Paired Variables** in the **Paired Samples T Test** dialog window.

- **Print** your output file to submit to your instructor.
- Answer the following questions for the affiliation variable:
    - What is the mean task skills score before the workshop?
    - What is the mean task skills score after the workshop?
    - How many degrees of freedom are associated with this test?
    - What is the value of the t-statistic?
    - What is the significance level?
    - Should you reject or fail to reject the null hypothesis of no difference in task skills before and after the workshop?
- **Write an interpretation of the results** along with your answers to the above questions to submit to your instructor. Follow the example in Section 10.3.

Chapter 11
One-way ANOVA:
Differences Between Multiple Independent Group Means

11.1 Introduction to One-way ANOVA

In the previous two chapters we distinguished between two distinct applications of the *t*-test - the independent samples *t*-test and the correlated samples *t*-test. This same distinction applies to the new procedure introduced in this chapter, **One-way ANOVA**. The present chapter will introduce this procedure for independent samples, and the ANOVA procedure for correlated samples will be presented in Chapter 12.

As with the *t*-test, ANOVA also tests for significant differences between groups. But while the *t*-test is limited to the comparison of only *two* groups, one-way ANOVA can be used to test differences in **three or more groups**. Several hypotheses worth investigating in our project involve the comparison of more than two groups. For example, EZ execs would certainly be interested in any differences in leader performance in relation to employees' **sex role type** or **leadership style**. However, as you will see, there are four levels of both of these quasi-independent variables (and so four groups to compare). Since the *t*-test only permits comparison of two groups, it is necessary to use analysis of variance (ANOVA) procedures for these comparisons.

The ANOVA procedure produces an ***F*** statistic, a value whose probability enables the researcher to reject or retain the null hypothesis, i.e., to conclude whether or not the differences in the scores on the dependent variable are statistically significant or due to chance. As with the *t*-test test, ANOVA is appropriate when the data are interval or ratio level, when the groups show similar variances, and when the data are normally distributed.

ANOVA is based upon a comparison of variance attributable to the independent variable (variability **between** groups or conditions) relative to the variance **within** groups resulting from random chance. In fact, the formula for computing the *F* statistic involves dividing the between-group variance estimate by the within-group variance estimate.

While different procedures are used to compute the ANOVA for independent vs. correlated samples designs, each procedure yields an *F* statistic which is evaluated in essentially the same way. Thus, regardless of which design is employed, the end result is an *F*-value, and when the probability of occurrence of the *F* value is less than .05, we conclude that there are significant differences groups, i.e., variation which cannot be attributed to chance.

11.1 a The Need for Individual Comparisons

When three or more groups are being analyzed in the ANOVA, there frequently arises the need to carry out more specific two-group comparisons in order to determine where the major treatment effect is occurring. A significant *F*-value only indicates that there is a significant difference *somewhere* between the groups - it does not indicate *which* groups are different. To determine this, secondary comparisons among all group means are needed subsequent to the ANOVA.

These two-group comparisons are commonly referred to as individual comparisons, follow-up tests, or *post-hoc* tests. Let's assume, for example, that you are a clinical psychologist who has designed an experiment to assess the effectiveness of different types of therapy on the treatment of phobias in a sample of eighty men and women. You assign each participant to one of four groups (either a control group or one of three different types of therapy). After an adequate treatment period, you rate the amount of improvement each subject has shown (higher score = more improvement). The mean improvement scores (hypothetical, of course!) for the four groups are as follows:

- Group 1 Control: $M = 12$
- Group 2 Psychoanalytic: $M = 18$
- Group 3 Client-centered: $M = 23$
- Group 4 Cognitive-behavioral: $M = 41$

You perform an ANOVA on the four groups, and find that there is significant overall variation (or differences) between the treatment groups. From this analysis, you can conclude that the independent variable (therapy) is having a significant effect, but you are unable to state where the effect is occurring. Is psychoanalytic therapy more effective than receiving no treatment at all? Is cognitive-behavioral therapy more effective than either psychoanalytic or client-centered therapies? Might there also be a significant difference between psychoanalytic and client-centered approaches? These questions require that we carry out more specific individual comparisons between the pairs of groups in which we're interested.

There are a number of guidelines for performing these two-group comparisons. For example, most researchers agree that it is appropriate to perform individual comparisons only if the result of the overall ANOVA is significant. Furthermore, one must decide at some point in the experimental process whether these comparisons are to be *a priori* (planned before the data are collected), or *post hoc* (decided upon after collecting and studying the data). Last, one must decide which specific *a priori* or *post hoc* technique (since there are many) best suits the situation (or researcher). While we are not able to review all the relevant guidelines for performing individual comparisons within this text, suffice it to say that these comparisons are generally considered an important and necessary part of the analysis of an experimental design in which there are three groups or more.

11.2 Sex Role Type and Leader Performance

For this chapter's example we will examine differences in **leader performance** (the dependent variable) in relation to EZ employees' **sex role type** (the independent variable). Recall from Chapter 5 that we referred to sex role type in the context of creating new variables using the SPSS **Transform** procedure, but we did not describe the specific rules that we used for creating this new variable. Essentially, we used a combination of scores on two categorical variables, **masc** and **fem**, to create the four levels of this new variable. That is, employees were classified into one of four sex-role identities based on their scores on **masc** and **fem**: **Sex-role diffuse**, **Masculine sex-typed**, **Feminine sex-typed** and **Androgynous**.

Specifically, employees were assigned a score between **1** and **4** on this new **sex role type** variable based on the *particular combination* of **masc** and **fem** scores each participant received. Thus, for any given participant, the combination of his/her scores on these two variables yields one of four possible categories into which s/he can be classified:

- low masculinity-low femininity (i.e., the employee scored 1 on both **masc** and **fem**)
- high masculinity-low femininity (the employee scored 2 on **masc** and 1 on **fem**)
- low masculinity-high femininity (the employee scored 1 on **masc** and 2 on **fem**)
- high masculinity-high femininity (the employee scored 2 on **masc** and 2 on **fem**)

To create these four groups, the SPSS **Transform, Into a Different Variable...** option was selected, then the following four conditional **IF** statements were used to create the new **sex role type** variable:

- If **masc** = 1 *and* **fem** = 1, then **sex role type** = **1** (Sex Role Diffuse Identity)
- If **masc** = 2 *and* **fem** = 1, then **sex role type** = **2** (Masculine Sex-typed Identity)
- If **masc** = 1 *and* **fem** = 2, then **sex role type** = **3** (Feminine Sex-typed Identity)
- If **masc** = 2 *and* **fem** = 2, then **sex role type** = **4** (Androgynous Sex Role Identity)

SPSS processed these conditional statements by examining each participant's **masc** and **fem** scores. Employees who received a score of **1** on *both* **masc** and **fem** (i.e., employees who were *both* low-masculine *and* low-feminine) were assigned the number **1** on the new **sex role type** variable (recall that this category of sex-role orientation is referred to as **Sex role diffuse**, indicating that persons in this group rate themselves as *neither* strongly masculine nor strongly feminine).

Employees who scored 2 on **masc** (high-masculine) *and* 1 on **fem** (low-feminine) were assigned the number **2** on **sex role type** (this represents the sex-role category, **Masculine sex-typed**, indicating that these people rate themselves as high-masculine and low-feminine). Next, employees who scored **1** on **masc** (low-masculine) *and* **2** on **fem** (high-feminine) were assigned the number **3** on **sex role**

type (this is the **Feminine sex-typed** sex-role category, consisting of people who rate themselves as low-masculine and high-feminine).

Finally, employees who received a score of **2** on *both* **masc** and **fem** (i.e., persons who scored high on both masculinity and femininity) were assigned the number **4** on **sex role type**(recall that this sex-role category is referred to as **Androgynous**, indicating that these people rate themselves as *both* strongly masculine and strongly feminine).

To see this new variable and the labels we assigned, open your **ezdata.sav** file and click the **Variable View** tab at the bottom of the window. Scroll to the bottom and you will see that we named this variable **sextype** in the data file. Go across that row and click on the cell in the **Values** column and you will see the labels we provided four the scores between 1 and 4 on this variable. In fact, you will also see another new variable, **leaderstyle**, and its labels - a variable we created using the same procedures. This will be the independent variable in the exercise at the end of this chapter, so we will describe it then.

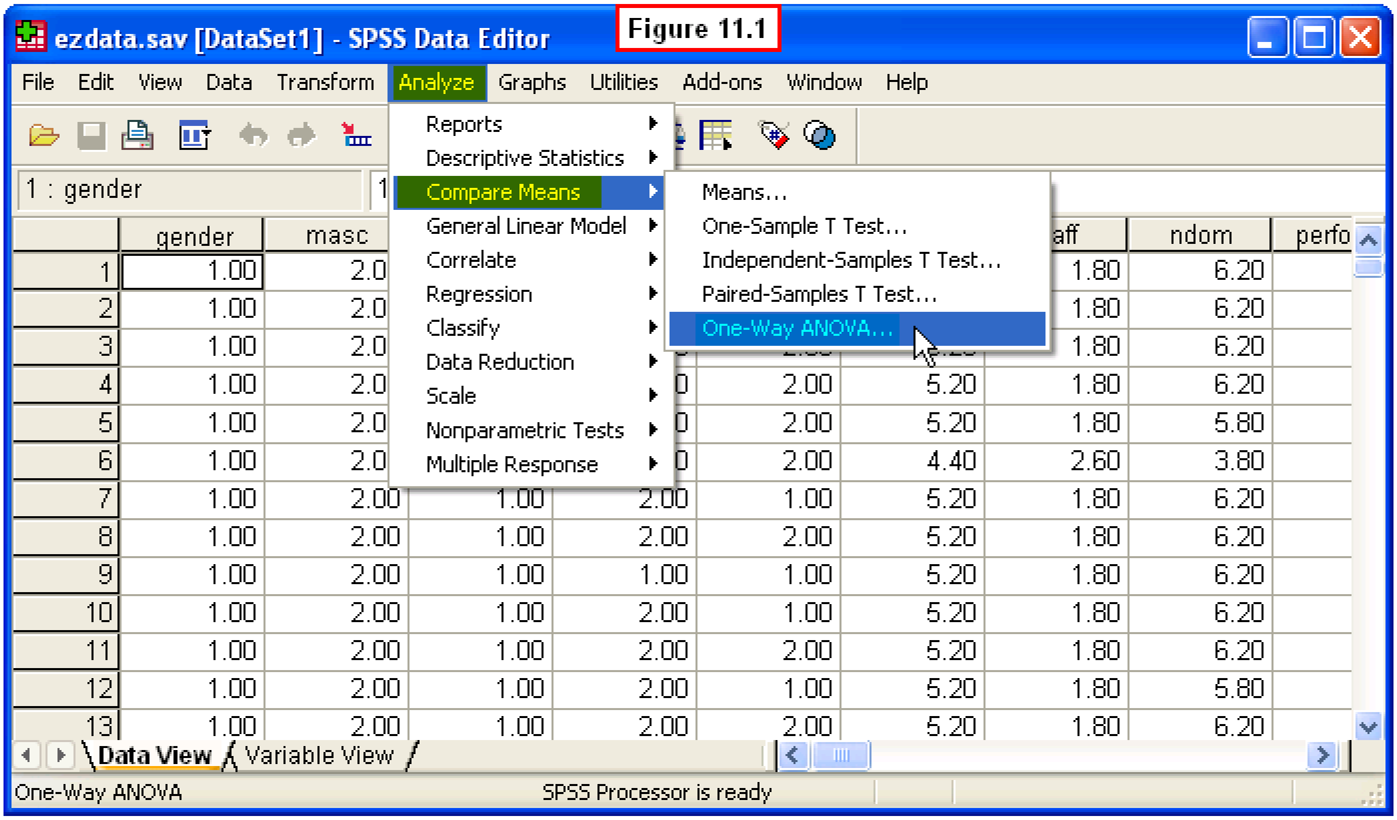
This new variable may seem confusing to you because of the complexity of the variable transformation procedure to create this new variable. In fact, that is why we chose not to include the description of this process in Chapter 5, and created the new variable for you! Also, the conceptual underpinnings of **sex role type** can be difficult to understand given that one must think of two variables simultaneously (**masc** and **fem**) to describe a person's sex role identity. But SPSS has already created the four groups for you, so just focus on the meaning of the four categories:

- **Sex role diffuse** persons see themselves as *neither* strongly masculine nor strongly feminine. They do not have a clearly developed identity regarding masculinity and femininity.
- **Masculine sex-typed** persons see themselves as primarily masculine in their identity. They emphasize masculinity and reject femininity in their self-concept.
- **Feminine sex-typed** persons see themselves as primarily feminine in their identity. The emphasize femininity and reject masculinity in their self-concept.
- **Androgynous** persons have integrated masculinity and femininity in their identity, seeing themselves as *both* strongly feminine *and* strongly masculine.

Given these four different types of sex role identity, our goal is to examine differences in our dependent variable, **leader performance**, in relation to this independent variable using the **One-way ANOVA** procedure. You might try to anticipate the results before we run the procedure. For example, one might expect that the sex role diffuse employees will have the lowest performance scores, while androgynous individuals will have the highest performance scores.

11.3 Running One-way ANOVA

Open your **ezdata.sav** file and select **Analyze, Compare Means, One-Way ANOVA...** from the menu (Figure 11.1).



A One-Way ANOVA dialog window will appear. Select **Leader Performance 1 (perform1)** on the left and move it to the **Dependent List:** pane on the right. Next scroll all the way down on the left, select **Sex Role Type (sextype)** and move it to the **Factor:** pane on the lower right. Note that SPSS uses the term, **Factor**, to refer to the independent variable, so these terms are synonymous. This will generate the ANOVA table when we run the analysis. But recall that we need to specify the multiple comparisons of group means first, so to do this click the **Post Hoc...** button at the bottom.

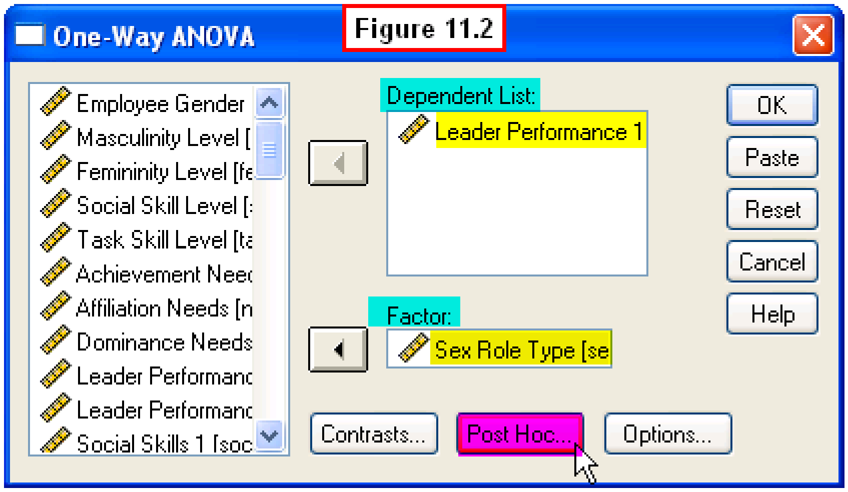


Figure 11.2

A **One-Way ANOVA: Post Hoc Multiple Comparisons** window will appear (Figure 11.3). There are numerous statistical procedures one can choose from to perform these comparisons, but a discussion of these is beyond the scope of this chapter. For our purposes, select the **LSD** box. LSD stands for Least Significant Difference.

Briefly, a value is computed which is how much two means must differ to justify rejecting Ho and concluding they are significantly different. The differences between all possible pairs of means are computed by subtraction. These mean differences are compared to the Least Significant Difference value. Any two means that differ at least by this amount will be judged to be significantly different at the .05 level.

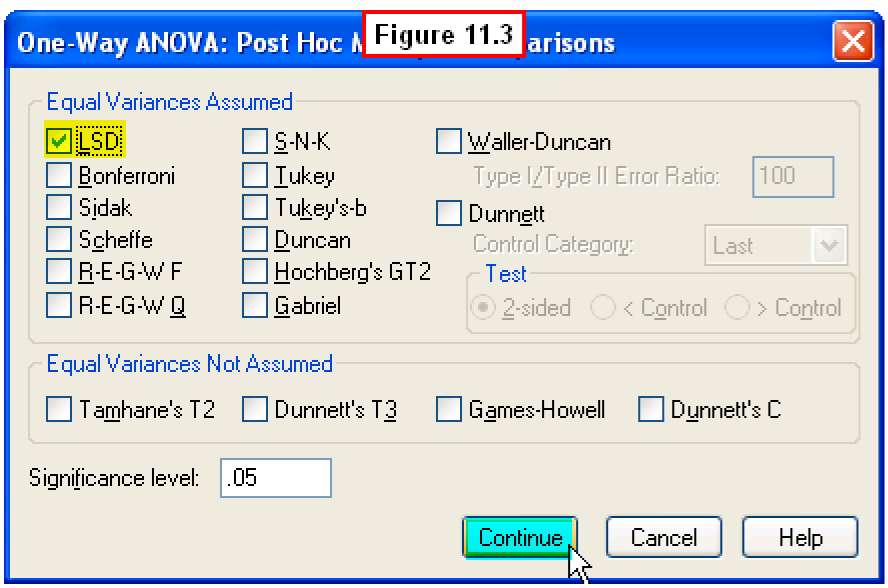
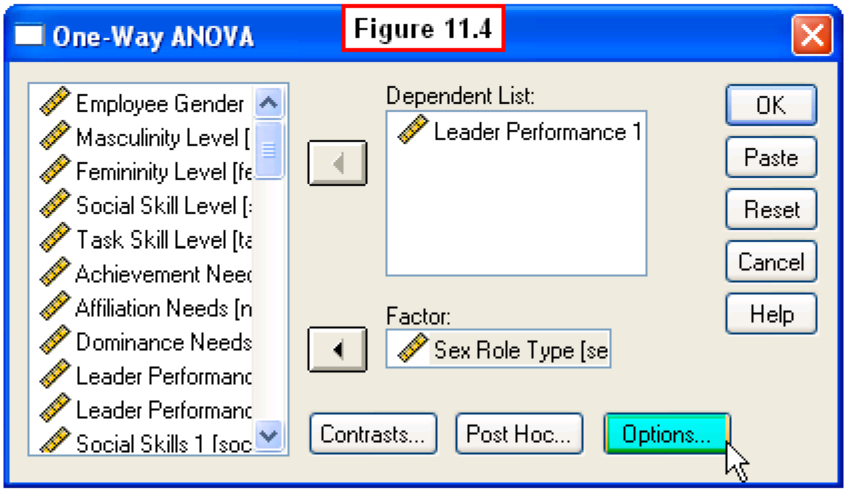
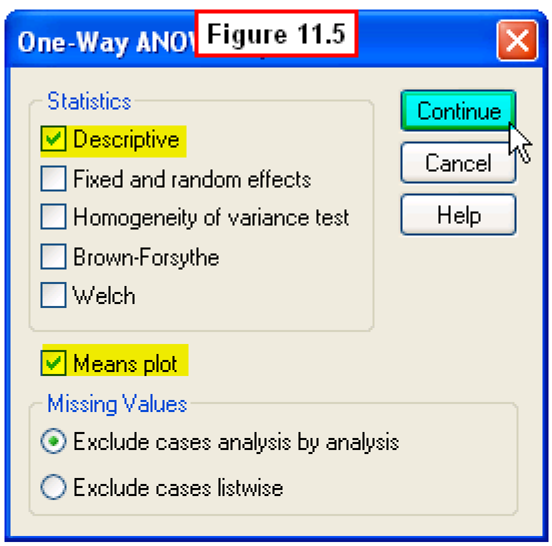


Figure 11.3

Click the **Continue** button at the bottom of this window and SPSS will return to the original **One-Way ANOVA** dialog window. Next we need to specify what other statistics besides the *F*-value that we want to compute. To do this, click the **Options...** button at the bottom of this window (Figure 11.4).



A **One-Way ANOVA: Options** dialog window will appear (Figure 11.5). Researchers typically want to examine descriptive statistics, so check the **Descriptive** box. Researchers also typically select the Homogeneity of variance test as well to determine whether or not this condition of ANOVA has been met. However, to simplify this example, we will not perform this test here. Last, recall that a visual depiction of results can greatly facilitate interpretation, so check the **Means plot** box. Then click the **Continue** button.



This will return you to the main dialog window. To run the ANOVA, click the **OK** button in the upper right corner.

11.4 Interpreting the Output

Several statistics are presented in the first table, **Descriptives** (Figure 11.6). The most relevant for our purposes are the four group means reflecting leader performance scores. Examination of these means suggests that the average leadership performance was lowest among the sex role diffuse group and highest for the androgynous group. But before we can interpret these means, we must first examine the results of the *F*-test and the multiple comparisons of means.

Figure 11.6

Descriptives

Leader Performance 1

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maxim
					Lower Bound	Upper Bound		
Sex Role Diffuse	37	3.4054	1.67431	.27525	2.8472	3.9636	1.00	[illegible]
Masc Sex-Typed	59	6.3898	2.31946	.30197	5.7854	6.9943	1.00	[illegible]
Fem Sex-Typed	51	5.0588	1.85916	.26033	4.5359	5.5817	2.00	[illegible]
Androgynous	81	7.2716	1.78194	.19799	6.8776	7.6656	2.00	[illegible]
Total	228	5.9211	2.36772	.15681	5.6121	6.2300	1.00	[illegible]

The ANOVA table is shown in Figure 11.7. The Mean Square values are computed by dividing the Between and Within Groups Sum of Squares by their respective degrees of freedom (df), where df = 3 and 224, respectively. The *F*-value (38.48) is computed by dividing the Mean Square Between Groups by the Mean Square Within Groups. As with the *t*-test, the most important part of this table the **Sig.** value, since this is the probability that the differences between groups is due to chance. Recall that if *p* (Sig.) is less than or equal to .05, we reject Ho. Since in this case, the Sig. is less than .001, we can reject the null hypothesis of no differences. If we were to write these results according to APA guidelines, we would state the following:

A one-way ANOVA revealed that there were significant differences in leadership performance between the four groups, $F\ (3,224) = 38.48$, $p < .001$.

Figure 11.7

ANOVA

Leader Performance 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	432.778	3	144.259	38.478	.000
Within Groups	839.801	224	3.749		
Total	1272.579	227			

Before we can state where the significant differences occur, we must examine the results of the LSD multiple comparisons (Figure 11.8). This table presents the results of all possible comparisons between means, so some of the information in the table is redundant. We will not discuss the section labeled **95% Confidence Interval**, since our main interest in this table lies with the results of the multiple comparisons.

Figure 11.8

Multiple Comparisons

Dependent Variable: Leader Performance 1

LSD

(I) Sex Role Type	(J) Sex Role Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Sex Role Diffuse	Masc Sex-Typed	-2.98443*	.40604	.000	-3.7846	-2.1843
	Fem Sex-Typed	-1.65342*	.41814	.000	-2.4774	-.8294
	Androgynous	-3.86620*	.38420	.000	-4.6233	-3.1091
Masc Sex-Typed	Sex Role Diffuse	2.98443*	.40604	.000	2.1843	3.7846
	Fem Sex-Typed	1.33101*	.37021	.000	.6015	2.0605
	Androgynous	-.88177*	.33141	.008	-1.5348	-.2287
Fem Sex-Typed	Sex Role Diffuse	1.65342*	.41814	.000	.8294	2.4774
	Masc Sex-Typed	-1.33101*	.37021	.000	-2.0605	-.6015
	Androgynous	-2.21278*	.34612	.000	-2.8948	-1.5307
Androgynous	Sex Role Diffuse	3.86620*	.38420	.000	3.1091	4.6233
	Masc Sex-Typed	.88177*	.33141	.008	.2287	1.5348
	Fem Sex-Typed	2.21278*	.34612	.000	1.5307	2.8948

*. The mean difference is significant at the .05 level.

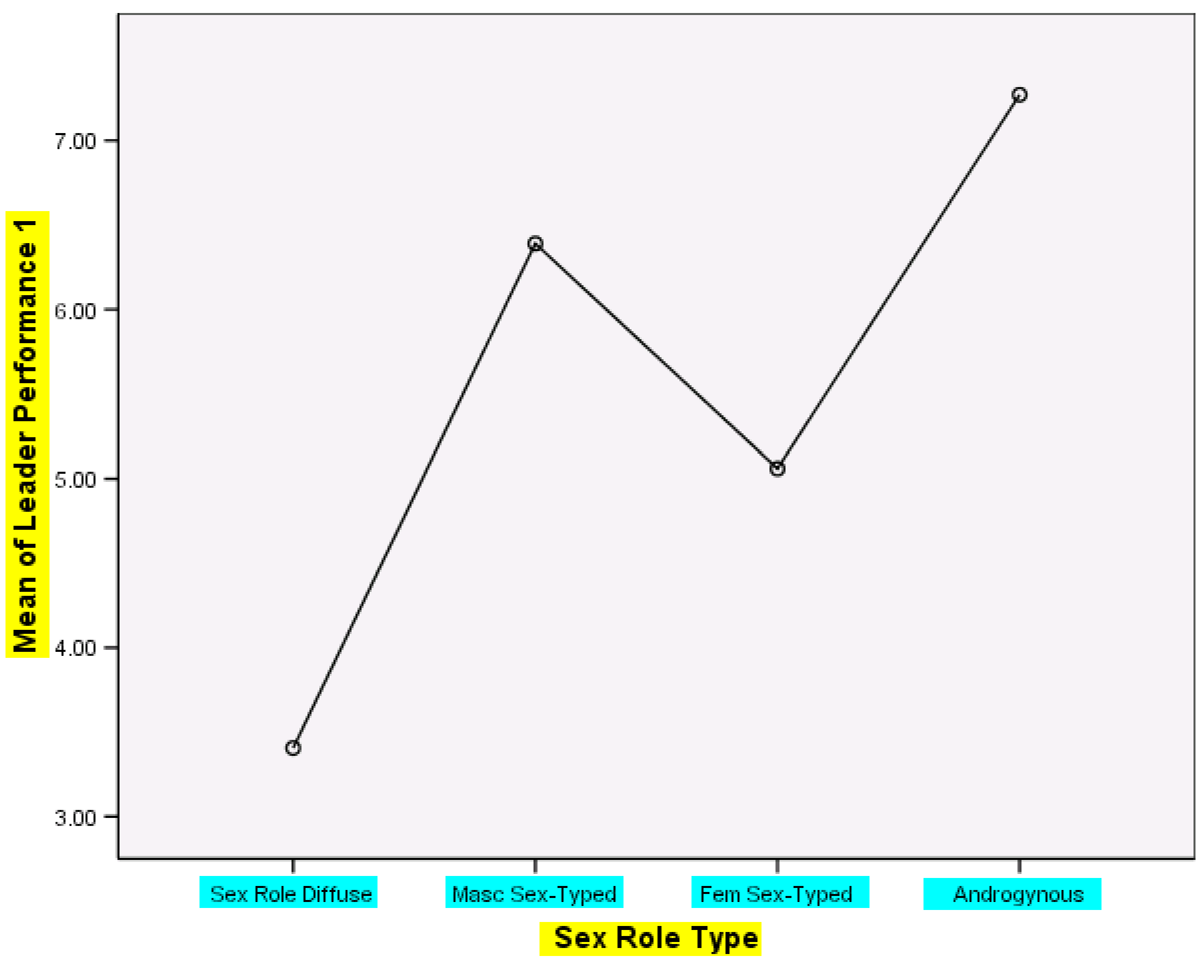
This table is divided into four main sections, with four rows within each section. For example, the first main section lists the Sex Role Diffuse group in the first column. The mean performance score for this group will be compared to that of the other three groups listed in the second column. In each main section, (**I**) refers to the mean for the group in the first column (e.g., Sex Role Diffuse), and (**J**) refers to the means for the other three groups listed in the second column (e.g., Masculine, Feminine, and Androgynous) which are being compared to the (**I**) mean.

The Mean Difference (**I - J**) column reports the difference between each pair of means. Look back at Figure 11.6 to see the group means. Thus, in the first row of this section, we can see that SPSS subtracted the Masculine sex-typed mean (**J** = 6.39) from the Sex role diffuse mean (**I** = 3.41), to yield (**I - J**) = (3.41 - 6.39) = -2.98 for the mean difference between these two groups. An asterisk next to the Mean Difference flags the pair of group means as significantly different at the .05 level (note that wherever an asterisk appears, the probability is less than .05 in the **Sig.** column for that particular comparison of two means). Below is the interpretation of these analyses as it would be written in a research report.

LSD comparisons revealed that all four means were significantly different from each other. Androgynous employees had significantly higher mean leadership performance ratings (M = 7.27) than did the masculine sex-typed (M = 6.39), feminine sex-typed (M = 4.88), and sex role diffuse employees (M = 3.41). Masculine sex-typed employees received significantly higher performance ratings than did the feminine sex-typed and sex role diffuse employees. Feminine sex-typed employees received higher performance ratings than did sex role diffuse employees.

These differences can be clearly seen in the graph depicting the mean performance scores of each of the four groups (Figure 11.9).

Figure 11.9



The one-way ANOVA procedure we performed yielded interesting results that you would want to report to upper-level management at EZ Manufacturing. Androgynous individuals (who integrate masculinity and femininity in their identity) appear to have the greatest leadership potential. Masculine sex-typed individuals did have higher performance scores than did feminine sex-typed employees. Perhaps this was due to a masculinity bias, reflecting stereotypes that associate masculinity, but not femininity, with leadership.

However, femininity was not viewed as a liability *per se*. Feminine sex-typed employees had higher performance scores than did the sex role diffuse group. Further, high-femininity, when paired with high-masculinity (i.e., androgyny) was associated with the best leader performance. These results demonstrate the importance of high masculinity in this organization, but also suggest that high-femininity does not necessarily detract from leader performance.

11.5 Chapter Review Video

Review Me!

11.6 Try It! Exercises

1. Using One-Way ANOVA: Leadership Style and Leader Performance

We have created another new variable, **Leadership Style**, in the **ezdata.sav** file using the same procedure we described in Section 11.2 to create the sex role typing variable. The leader style variable was created using the categorical variables, **task1** (where 1 = low task skills; 2 = high task skills), and **soc1** (where 1 = low social skills; 2 = high social skills).

Specifically, employees were assigned a score between **1** and **4** on this new **leaderstyle** variable based on the *particular combination* of **task1** and **soc1** scores each participant received. Thus, for any given participant, the combination of his/her scores on these two variables yields one of four possible categories into which s/he can be classified:

- low task skills-low social skills (i.e., the employee scored 1 on both **task1** and **soc1**)
- high task skills-low social skills (the employee scored 2 on **task1** and 1 on **soc1**)
- low task skills-high social skills (the employee scored 1 on **task1** and 2 on **soc1**)
- high task skills-high social skills (the employee scored 2 on **task1** and 2 on **soc1**)

To create these four groups, the SPSS **Transform, Into a Different Variable...** option was selected, then the following four conditional **IF** statements were used to create the new **leaderstyle** variable:

- If **task1** = 1 *and* **soc1** = 1, then **leaderstyle** = **1** (Low Task & Social Skills)
- If **task1** = 2 *and* **soc1** = 1, then **leaderstyle** = **2** (High Task/Low Social Skills)
- If **task1** = 1 *and* **soc1** = 2, then **leaderstyle** = **3** (Low Task/High Social Skills)
- If **task1** = 2 *and* **soc1** = 2, then **leaderstyle** = **4** (High Task & Social Skills)

Thus, Group 1 consists of employees aren't oriented to either task or social issues in leadership style. Group 2 consists of people who are primarily task-oriented in their leadership style. Group 3 includes persons who are primarily relationship-oriented in their leadership style. Last, Group 4 consists of individuals who are oriented strongly towards both task and social issues in leadership style. You might try to anticipate the pattern of results regarding performance before you run this analysis.

Open your **ezdata.sav** file and you will see that this new variable is the last one listed in the data file. Use the procedures described in Section 11.3 to conduct a one-way ANOVA using **Leadership Style (leader style)** as the independent variable (factor) and **Leadership Performance (perform1)** as the dependent variable to determine whether or not there are significant differences in performance in relation to employee leadership style.

- **Print** your output file to submit to your instructor.
- Answer the following questions:
    - What is the value of the *F*-statistic?
    - What is the significance level?
    - Should you reject or fail to reject the null hypothesis of no differences in performance between the leader style groups?
    - Write a statement about the *F*-value as shown in Section 11.4
    - Conduct LSD Multiple Comparisons in your analysis. Based on the results of the post hoc tests, which groups differ from each other?
    - For those means that are significantly different, state what the differences are (i.e., which group had the higher performance score?)
- **Write an interpretation of the results** along with your answers to the above questions to submit to your instructor. Follow the example in Section 11.4.

Chapter 12

Repeated Measures One-way ANOVA: Differences Between Multiple Correlated Group Means

Raymond O'Connor, Jr.
Valparaiso University

12.1 Introduction to Repeated Measures One-way ANOVA

As in Chapter 11, the topic of this chapter is also One-Way ANOVA. Thus, much of what you learned in Chapter 11 will be applicable to this chapter. The major difference is that whereas we introduced **between groups** independent samples ANOVA in Chapter 11, in the present chapter we will discuss **within group** correlated samples ANOVA.

You may recall that one instance of a within groups design involves repeated measures on the same participants. The **before-after** design introduced in Chapter 10 is an example of this type of design. Recall that EZ employees were measured on performance, social skills and task skills before they attended a leadership workshop (**soc, task,** and **perform**), then they were re-measured on these variables immediately after the workshop (**soc2, task2,** and **perform2**).

In Chapter 10, we introduced the **paired-samples t-test** as the appropriate analysis for comparing scores before and immediately after the workshop. However, you may recall from Chapter 4 that we also included a **third measurement** of these variables three months after the workshop. This *delayed post test* is not uncommon with program evaluation. That is, researchers are not only interested in immediate gains after participation in the workshop, they want to know whether or not the change is long-lasting.

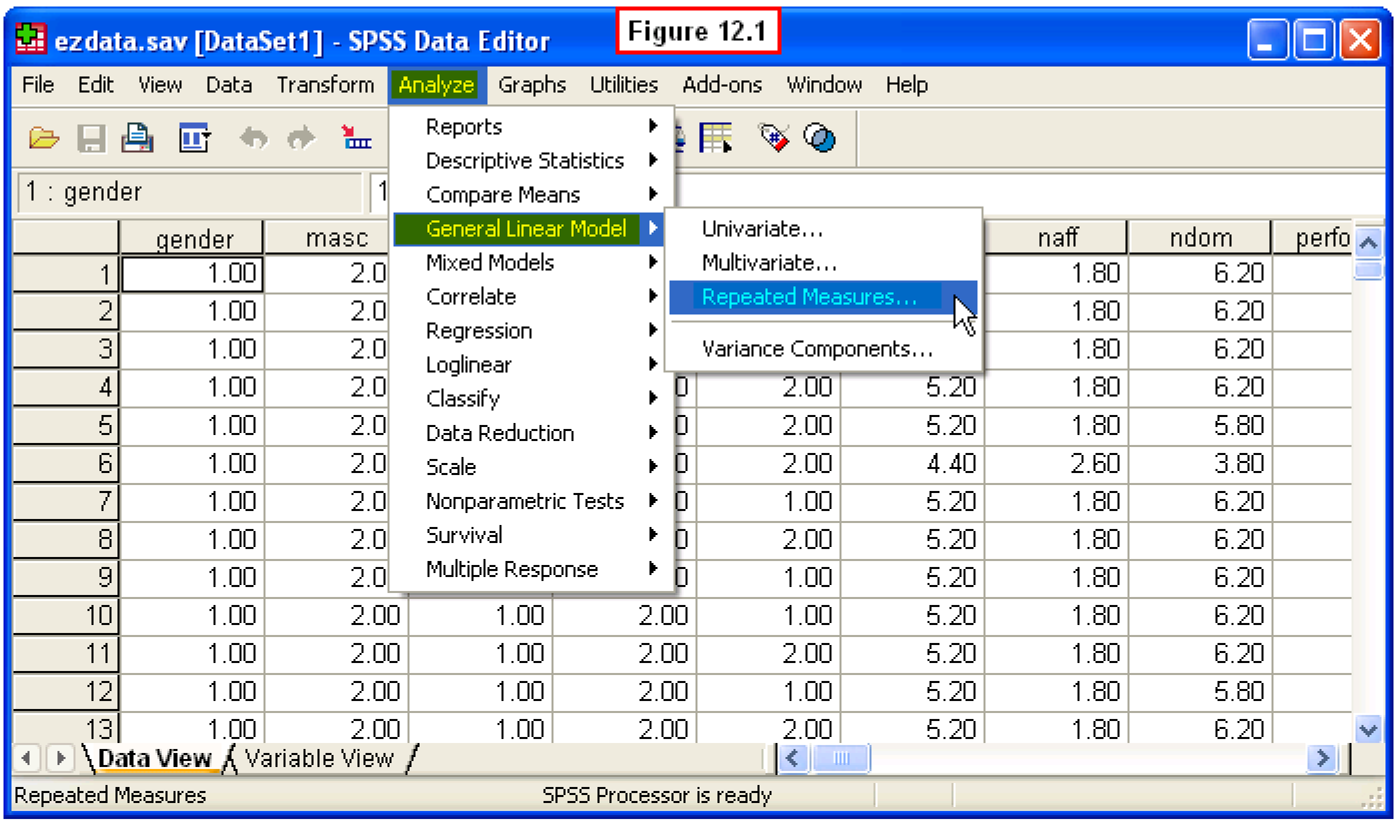
Thus, while we saw in Chapter 10 that there was a significant improvement in scores after the workshop, the focus in the present chapter is on whether or not that change was enduring, i.e., will the increases be maintained over three months? While the optimal situation would be to see the same level of improvement at three months, there is typically some loss over time. So many program evaluators consider it a success if the three-month post test shows that scores are still significantly higher than at pre-test, even if they may be lower than observed immediately after the workshop. The analyses introduced in this chapter will answer this question.

Thus, this design bears similarities to the two-group repeated measures t-test design in Chapter 10, but also shares characteristics of independent samples one-way ANOVA in that there are now three groups to be compared. So the appropriate analysis of this data is a **repeated measures one-way ANOVA**.

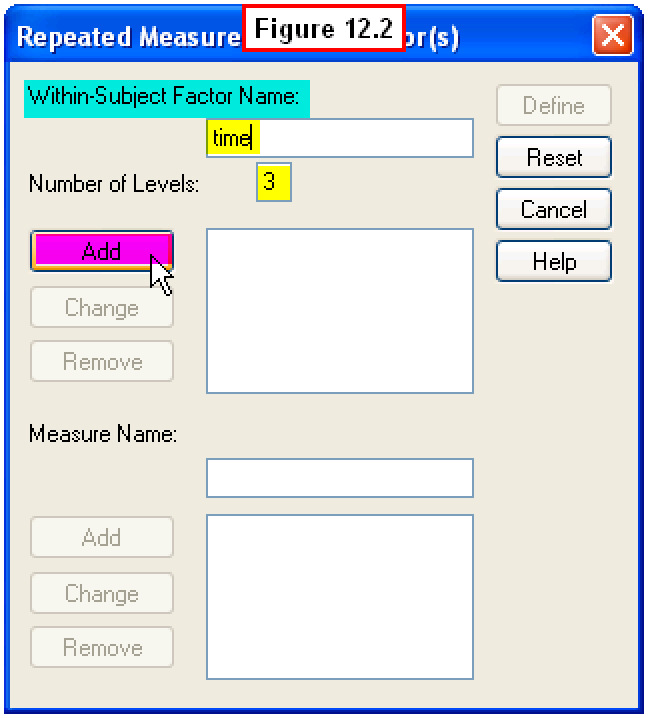
For the example of this design and analysis in this chapter, we will determine whether or not there are significant differences in the leadership performance scores at the three times of measurement (before, immediately after, and three months after the workshop). These variables are **perform, perform2,** and **perform3** in the ezdata file. For the chapter exercise, we will ask you to examine differences in the leader social skills scores across the three times (**soc, soc2,** and **soc3**).

12.2 Running Repeated Measures ANOVA

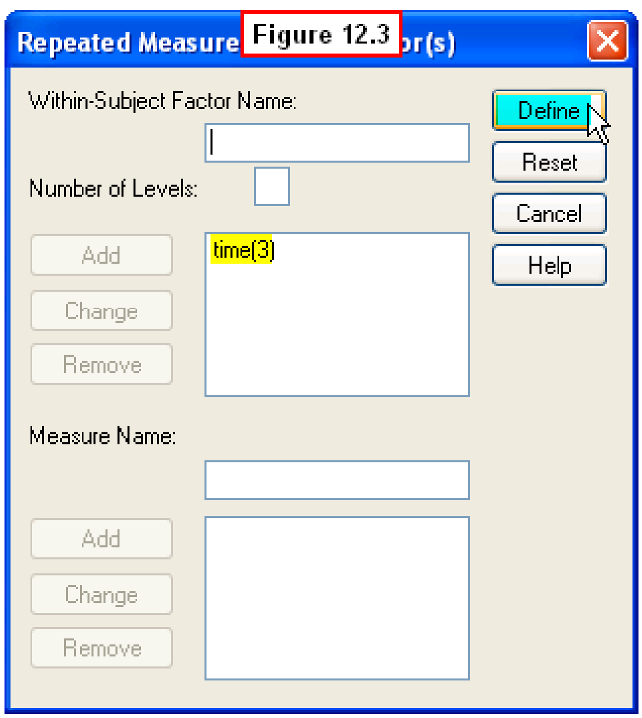
Open your **ezdata.sav** file and select **Analyze, General Linear Model, Repeated Measures...**from the menu (Figure 12.1).



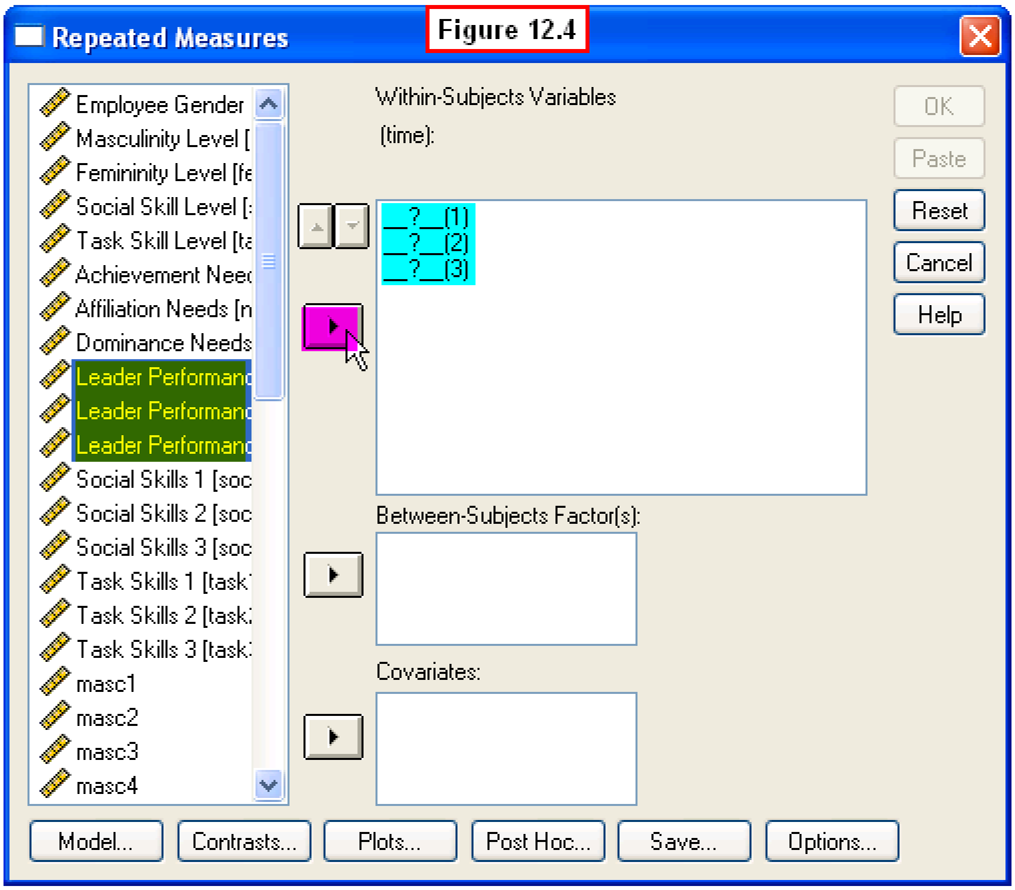
A **Repeated Measures Define Factor(s)** dialog window will appear (Figure 12.2). We need to define the **Within-Subject Factor Name** (i.e., the repeated measures variable, **perform, perform 2, perform3**), and enter the number of levels of this factor. To do this, type **time** in the box below **Within-Subject Factor Name**, and enter a **3** in the box next to **Number of levels**. Then click the **Add** button.



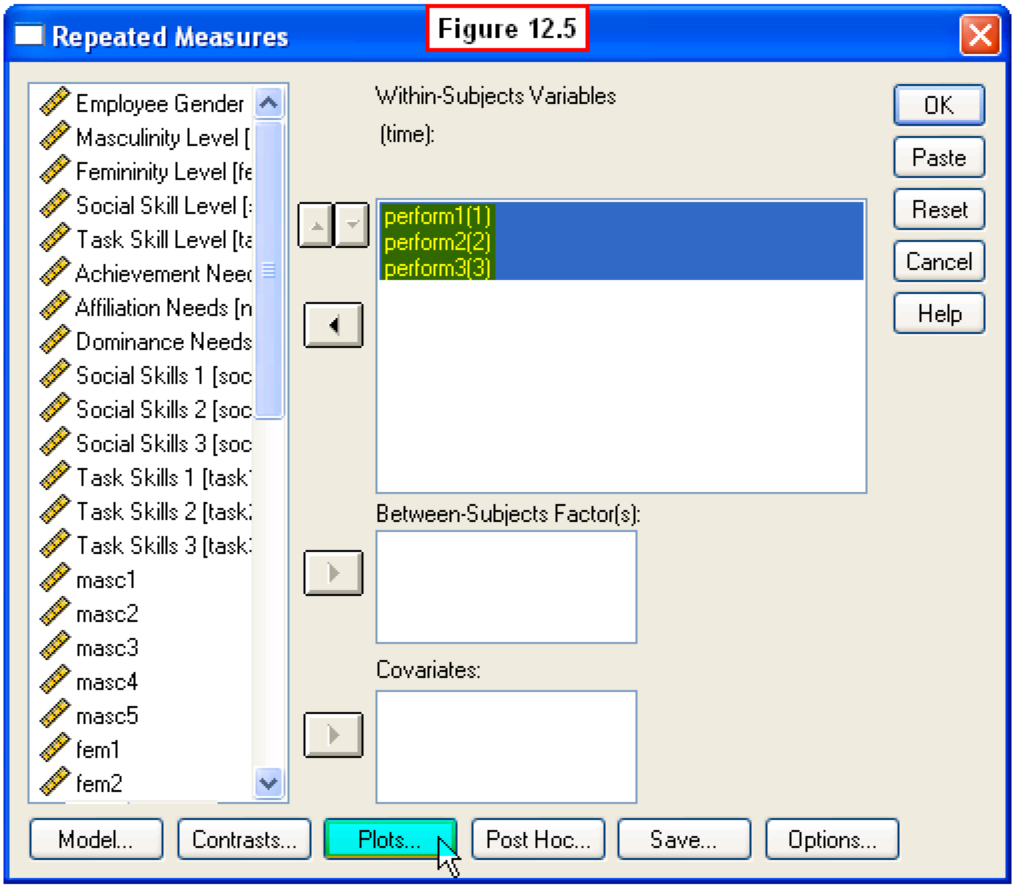
After you have done this, your window should look like that shown in Figure 12.3. Note that SPSS will now name this variable **time**, and it shows that there are three levels: **time (3)**. Now SPSS needs us to define what the three levels of time are. To do this, click the **Define** button in the upper right corner.



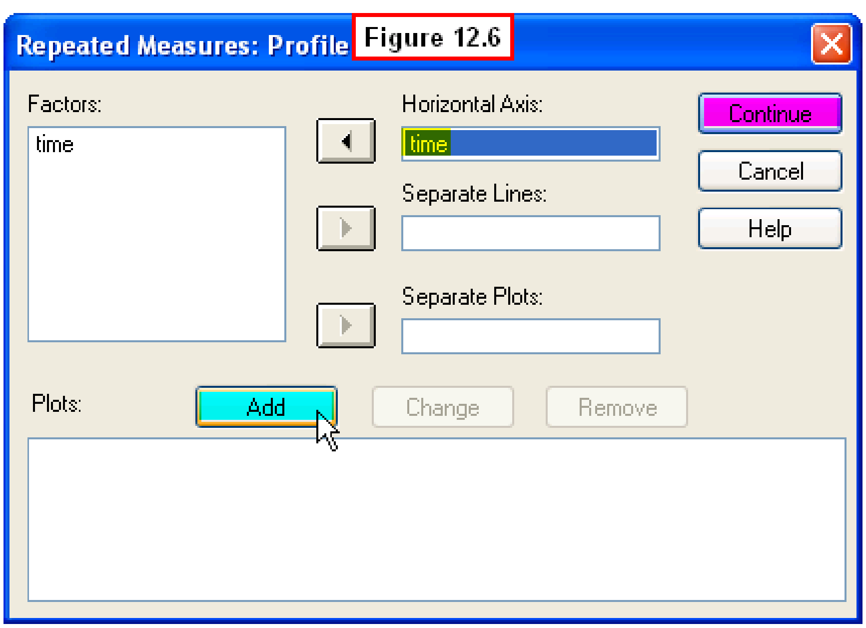
A **Repeated Measures** dialog window will appear (Figure 12.4). To specify the levels, scroll down the variable list on the left, then highlight the three **Leadership Performance** variables. Click the right-arrow button in the middle, and these variables will be moved to the appropriate places in the **Winthin-Subjects Variables (time)** variable definition box.



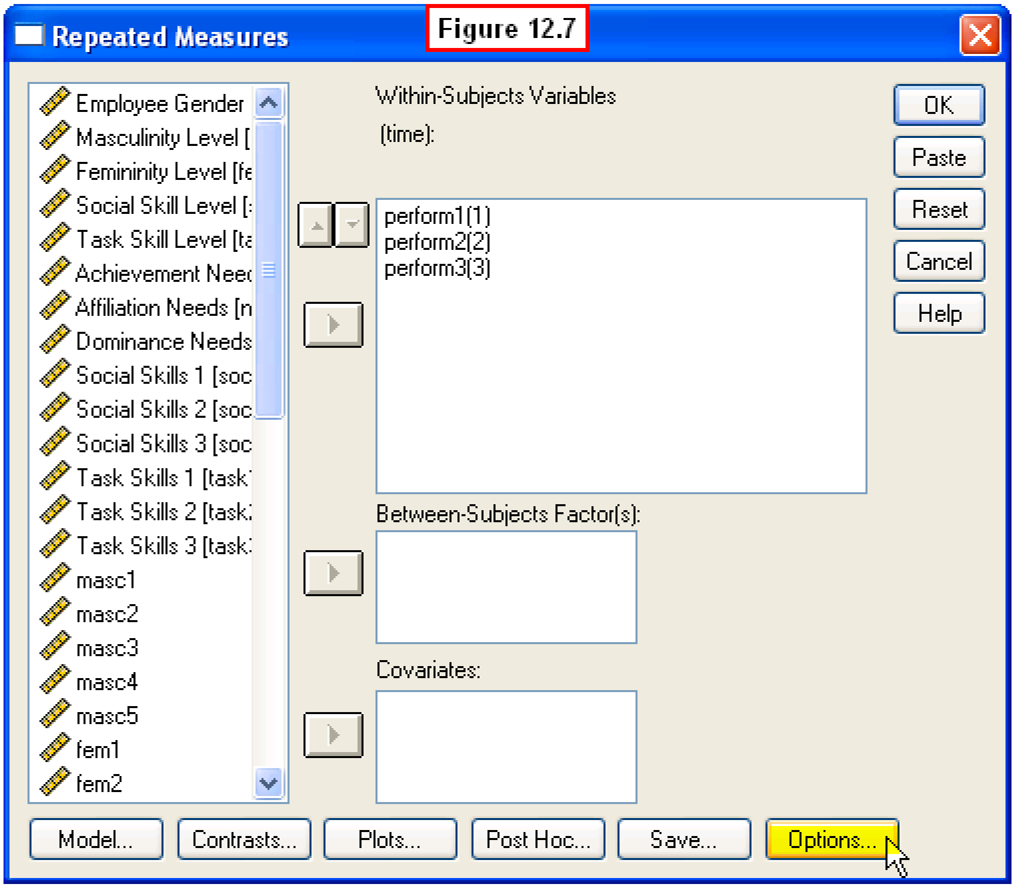
Your window should now look like the one shown in Figure 12.5. As we saw in Chapter 11, it is helpful to see a graph of the three performance score means, so to generate this, click the **Plots...** button at the bottom of this window.



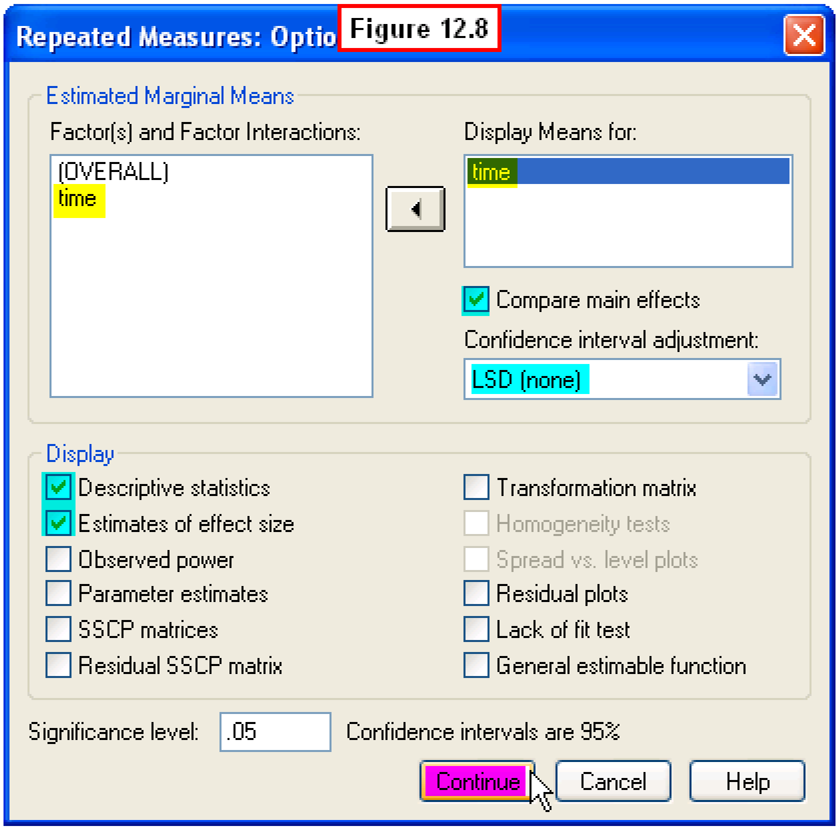
A **Repeated Measures: Profile Plots** dialog window will appear (Figure 12.6). Select the **time** variable on the left, then click the right-arrow button in the middle to move this to the **Horizontal Axis** box on the right. Then click the **Add** button at the bottom of this dialog window.



This will cause SPSS to create a figure displaying the mean performance scores on the Y-axis (vertical) and the three times of measurement on the X-axis (horizontal). Click the **Continue** button in the upper-right corner, and this will close this window and return you to the main **Repeated Measures** dialog window (Figure 12.7).



So far we have instructed SPSS to compute the repeated measures one-way ANOVA. Recall from Chapter 11 that the overall ANOVA results just tell us whether or not there is a significant difference *somewhere* between the three means. In order to determine which means are significantly different, we need to instruct SPSS to compute multiple comparisons of means following the ANOVA. To do this, click the **Options...** button in the lower-right corner. A **Repeated Measures: Options** dialog window will appear (Figure 12.8).



First highlight the **time** variable on the left and move it to the **Display Means for:** box on the right. Then check the **Compare main effects** box and select **LSD (none)** from the drop down menu in this box. Recall from Chapter 11 that the LSD test is a procedure for conducting multiple comparisons among the three mean performance scores.

Next, check the **Descriptive statistics** box to generate other descriptive statistics on these variables, and check the **Estimates of effect size** box to generate partial-eta square values to assess how large the differences between the means is (see Chapter 11). Last, click the **Continue** button at the bottom of this window.

This window will close and return you to the main **Repeated Measures** dialog window. To run our analysis, click the **OK** button in the upper right corner of this window. The results of this analysis will then be displayed in an output viewer window.

12.3 Interpreting the Output

The first table of the output simply repeats the definition of our within-groups factor, **time**. Next, the **Descriptive Statistics** table (Figure 12.9) presents the three performance means, their standard deviations and sample sizes. The most relevant statistics for our purposes are the three means reflecting leader performance scores.

Examination of these means suggests that the average leadership performance was lowest at time 1 (before the workshop), highest at time 2 (immediately after the workshop), and in between these two at time 3 (the 3-month follow-up measure). But before we can interpret these means, we must first examine the results of the *F*-test and the multiple comparisons of means. Also, skip over the next table on the output, **Multivariate Tests**, since it is beyond the scope of this chapter.

Figure 12.9

Descriptive Statistics

	Mean	Std. Deviation	N
Leader Performance 1	5.9211	2.36772	228
Leader Performance 2	6.9561	1.75047	228
Leader Performance 3	6.3816	1.80284	228

Before we examine the ANOVA results, we must first examine the **Mauchly's Test of Sphericity** (Figure 12.10). Complete discussion of this test is also beyond our scope, but interpretation of this statistic is similar to that of the Levine's Test for Equality of Variance (see Chapter 9, section 9.3a). Since the **Sig** probability (.000) is less than .05, we cannot assume that the variances between the three sets of scores are equal.

Figure 12.10

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.250	313.637	2	.000	.571	.572	.500

As we saw in Chapter 9, this means that a statistical correction needs to be made due to the lack of homogeneity (equality) of variance. SPSS provides this adjustment in the next table of the output (Figure 12.11). This table, **Tests of Within-Subjects Effects**, presents the ANOVA results. Note that there are four *F*-values shown in the first section of this table. The first row presents the results if sphericity could be assumed. The latter three values are different ways of statistically correcting for the lack of sphericity. The most conservative of the three is the **Lower-bound** test (which is highlighted in this table). In actuality, there was not much of an adjustment in that all four *F*-values are the same.

Figure 12.11

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	122.635	2	61.317	58.479	.000	.205
	Greenhouse-Geisser	122.635	1.143	107.328	58.479	.000	.205
	Huynh-Feldt	122.635	1.145	107.142	58.479	.000	.205
	Lower-bound	122.635	1.000	122.635	58.479	.000	.205
Error(time)	Sphericity Assumed	476.032	454	1.049			
	Greenhouse-Geisser	476.032	259.374	1.835			
	Huynh-Feldt	476.032	259.824	1.832			
	Lower-bound	476.032	227.000	2.097			

As we have seen in previous discussions of null hypothesis testing, since the **Sig.** level is less than .05, then we can reject the null hypothesis of no difference in performance between the three times of measurement in favor of the alternative hypothesis that there is a significant difference somewhere between the three performance means. The partial eta squared value (.205) indicates that this is a relatively small effect, even though it is significant.

Skip over the next table, **Tests of Within-Subject Contrasts**, since this analysis is not relevant here. Also skip over the next table of the output, **Tests of Between-Subjects Effects**, since there is no between-groups factor in this design. The next table, **Estimates**, again presents the three means, their standard errors and upper/lower bounds.

We are interested in comparing these means, so the next table, **Pairwise Comparisons**, is the one we need to examine. The multiple comparisons of the three means are presented in this table (Figure 12.12).

Figure 12.12

Pairwise Comparisons

Measure: MEASURE_1

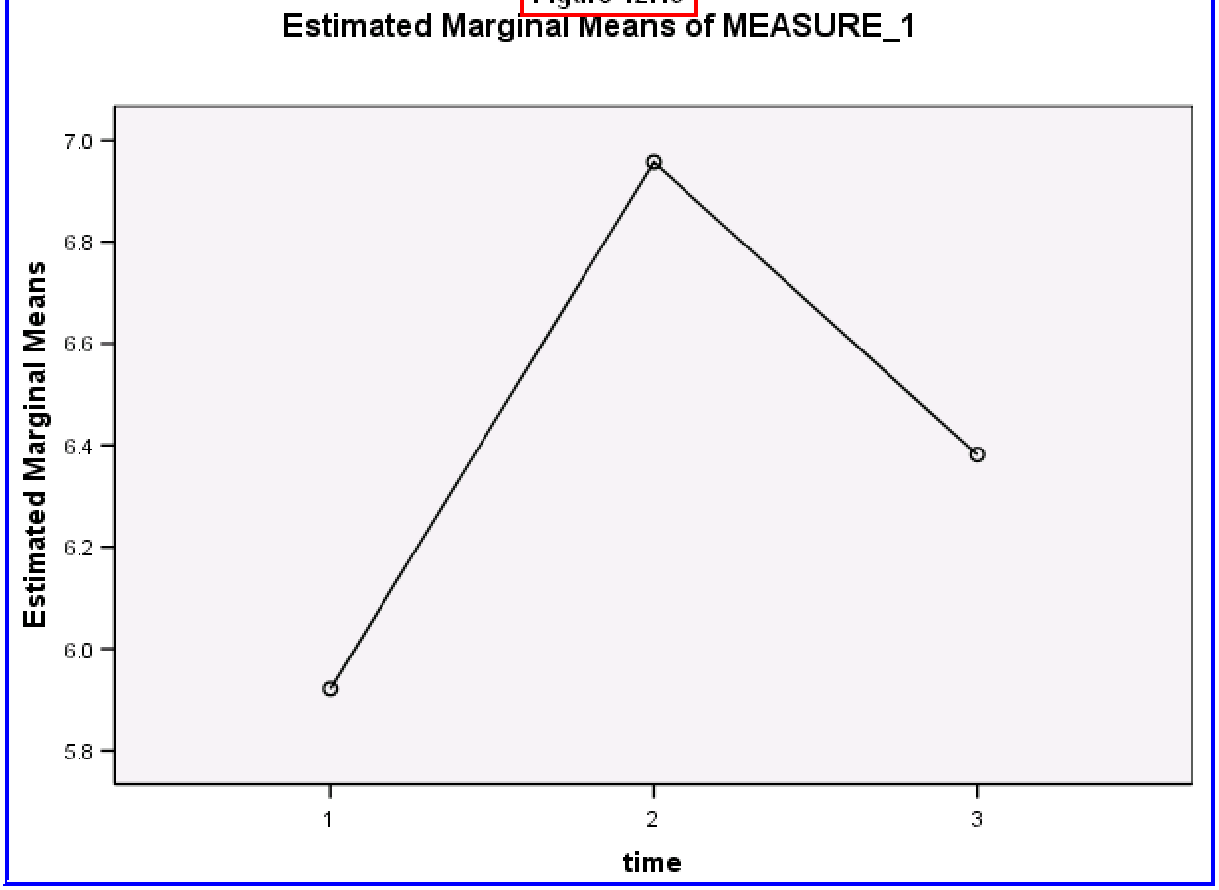
(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-1.035*	.126	.000	-1.282	-.788
	3	-.461*	.099	.000	-.656	-.265
2	1	1.035*	.126	.000	.788	1.282
	3	.575*	.045	.000	.486	.663
3	1	.461*	.099	.000	.265	.656
	2	-.575*	.045	.000	-.663	-.486

The first section of this table presents the comparisons of the mean performance scores at time 1 to those at time 2 and time 3. The second column shows the actual difference between each pair of means being compared. Our primary interest is in the **Sig.** column, because this indicates whether or not any given pair of means is significantly different. Since $p < .05$ in this column for both comparisons, we can conclude that performance at time 1 was significantly different from that at both time 2 and time 3. Note that this is also indicated by an asterisk placed next to the Mean Difference value.

Since we have already made two comparisons, all that remains is the comparison of the means at time 2 and time 3. This is shown in the second section of this table. The difference between time 2 and time 3 (i.e., .575) is statistically significant, since $p < .05$ for this comparison in the **Sig.** column. In sum, these comparisons reveal that all three means are significantly different from each other. The remainder of this table is redundant, since we have already discussed all three comparisons.

Scroll to the end of the output file to see the graph of these three means. The pattern of differences just described can be seen clearly in the plot of the three means shown in Figure 12.13.

Figure 12.13



This graph illustrates that performance increased significantly immediately after the workshop, but that performance was significantly lower three months later than immediately after the workshop. However, the performance at three months was still significantly higher than it was before the workshop. Below is an example of how these results would be written in APA format:

A repeated measures one-way ANOVA revealed that there were significant differences in leadership performance between the three times of measurement, $F(1,227) = 58.48$, $p < .001$, though this was a relatively small effect size (eta-squared = .21). LSD comparisons revealed that all three means were significantly different from each other. Mean leadership performance was significantly higher immediately after the leadership training workshop ($M = 6.96$) than before the workshop ($M = 5.92$). The mean performance score three months later was significantly lower ($M = 6.38$) than that immediately after the workshop, but significantly higher than the mean performance before the workshop.

The repeated measures one way ANOVA procedure we performed yielded interesting results that you would want to report to upper-level management at EZ Manufacturing. In particular, there is evidence that the leadership training program did significantly increase performance. While the largest increase was observed immediately after the workshop, and there was a significant decline after three month, the performance at three months was still significantly higher than before the workshop. This indicates that the workshop was effective at producing improved leadership, and that this increase was relatively enduring.

12.4 Chapter Review Video

[Review Me!]

12.5 Try It! Exercises

1. Using Repeated Measures One-Way ANOVA: Assessing Changes in Leader Social Skills

Use the procedures described in Section 12.2 to compute a repeated-measures one-way ANOVA to test for differences in leader social skills before, immediately after, and three months after EZ employees attended the leadership training workshop. This variable has three levels (**soc1**, **soc2** and **soc3**). Define this within-groups variable as **time**. Generate a plot of the means and request LSD comparisons by choosing the **Compare Main effects** option.

When you have generated your output file:

- **Print** your output file to submit to your instructor.
- Answer the following questions:
  - What is the value of the *F*-statistic?
  - What is the significance level?
  - Should you reject or fail to reject the null hypothesis of no differences in social skills across the three time periods?
  - Write a statement about the *F*-value as shown in Section 12.4
  - Based on the results of the LSD Multiple Comparisons tests (pairwise comparisons), which means are significantly different?
  - For those means that are significantly different, state what the differences are (i.e., state which social skills means are higher or lower than the others).
- **Write an interpretation of the results in APA format** to submit to your instructor. Follow the example in Section 12.4.

Chapter 13

Two-Way Factorial ANOVA: Using More than One Independent Variable

13.1 Introduction to Two-Way ANOVA

In the previous two chapters we introduced **One-Way ANOVA** designs that involve multiple levels of one independent variable (or factor). The present chapter will introduce a more sophisticated ANOVA model that allows the researcher to test the effectives of **two** independent variables; hence, this procedure is called **Two-Way ANOVA** (sometimes also called **Factorial ANOVA**). That is, factorial ANOVA improves on one-way ANOVA in that the researcher can simultaneously assess the effects of *two (or more) independent variables* on a single dependent variable within the same analysis. Thus, factorial ANOVA yields the same information that two one-way ANOVA's would, but it does so in one analysis.

But that's not all. Factorial ANOVA also allows the investigator to determine the possible *combined* effects of the independent variables. That is, it also assesses the ways in which these variables *interact* with one another to influence scores on the dependent variable. Although understanding such **interaction effects** can be a complex and difficult task, it is essential to the progress of science, since in the real world many variables interact with one another to determine behavior. In this chapter we provide a basic introduction to factorial ANOVA and the SPSS program that performs this powerful statistical analysis.

The conceptual basis of factorial ANOVA is essentially the same as that of one-way ANOVA, and the interpretation of the resulting F-values is also based on the same logic as in the one-way ANOVA. The difference is that where one-way ANOVA only generates one F-value, two-way ANOVA generates **three** F-values: one to test the **main effects** of each factor, and a third to test the **interaction effect** (i.e., the combined effect of the two factors).

13.1 a The need for a *factorial combination* of Independent Variables

A basic requirement for factorial experimental design is that the levels of the two independent variables have been completely crossed in a factorial combination. This means that each level of the first independent variable must be combined with each level of the other independent variable. In the simplest two-way ANOVA (a **2 x 2 design**), four different groups of participants would be needed. If the first factor (Factor A) is **GENDER** (where level **A1** is male employees and level **A2** is female employees), and the second factor (Factor B) is **MASC** (where level **B1** is low-masculine employees and level **B2** is high-masculine employees), four combinations would be required to permit a factorial ANOVA. Each unique combination is referred to as a **cell** (see Table

13.1). In this table, the four cells (in **boldface**) represent these four possible combinations.

Table 13.1			
		Gender (Factor A)	
		Males (A1)	Females (A2)
Masculinity (Factor B)	Low (B1)	**Low-masc Males (A1B1)**	**Low-masc Females (A2B1)**
	High (B2)	**High-masc Males (A1B2)**	**High-masc Females (A2B2)**

Thus, the four cells of the factorial combination shown in Table 13.1 for a **2 (Gender) x 2 (Masculinity)** factorial design would require us to divide our participants into the following four groups:

(1) Low-masculine males (A1B1)
(2) High-masculine males (A1B2)
(3) Low-masculine females (A2B1)
(4) High-masculine females (A2B2)

This factorial combination will allow us to compare the scores of men vs. women (**GENDER**) and low-masculine vs. high-masculine employees (**MASC**) on a given dependent variable. This would be the same as if we did two separate studies and conducted two t-tests (one comparing the male vs. female scores, and one comparing low- vs. high-masculine employees' scores). But it would be more economical and efficient, because we would get the same information from one study and one analysis (the 2 x 2 ANOVA). What is crucial to the factorial combination of these two independent variables is that we are also able to assess the possible interaction effect of the two independent variables combined.

13.2 Main Effects & Interactions

For this chapter's example we will examine differences in overall masculinity (i.e., **MASCTOT**, the dependent variable) in relation to EZ employees' **sex** **(GENDER)** and **masculinity category** **(MASC)**, the two independent variables. Recall from the two previous chapters that all ANOVA's are conceptually based upon a comparison of variance attributed to the independent variable (called *between-groups*, or *treatment* variance) to the variance due to random fluctuation (called *within-groups*, or *error* variance). Factorial ANOVA is conceptually based on the same type of ratio computations. The difference is that a 2 x 2 factorial ANOVA yields three F-ratios, instead of only one as in the one-way ANOVA. In the following we will briefly review the reasons for this and show how to interpret these three F-values.

13.2a Sources of Variance in Factorial ANOVA

The reason that **three F-ratios** are computed is that conducting factorial ANOVA is like performing two one-way ANOVA's to assess the simple **Main Effects** of each of the two independent variables (**Factors A** and **B**), and a third analysis to assess the **A x B interaction effect** on the dependent variable (**MASCTOT**). Specifically, in a between-groups design, between-groups variance can be separated into three sources:

- variance due to the **main effect of Factor A** (**GENDER**)
  - masculinity differences between the two groups of men vs. women
  - evaluated by comparing the *column* ***marginal means***
- variance due to the **main effect of Factor B** (**MASC**)
  - masculinity differences between the two groups of low vs. high-masculinity categories
  - evaluated by comparing the *row* ***marginal means***
- variance due to the **A x B interaction effect** (**GENDER x MASC**)
  - masculinity differences between the four groups of combinations of low and high masculine men and women
  - evaluated by comparing the ***cell means*** of the factorial combination.

Three separate F-ratios are computed to determine how much of the variance in the dependent variable can be attributed to each of these three effects. Each F-value represents the ratio of the variance from that particular effect relative to random error variance. Thus, the three sources of between-group variance (Factor A main effect, Factor B main effect, and A x B interaction) result in three F-values. As in the one-way ANOVA, the significance of each effect is decided by looking at the probability associated with each F-value (i.e., if $p < .05$, the effect is significant).

13.2 b Interpreting the Results of Factorial ANOVA

The **interpretation of main effects** from a 2 x 2 factorial ANOVA is straightforward. If $p < .05$ for the main effect of a particular factor, then there is a significant effect for that factor. All we have to do is examine the marginal means for the levels of the factor to determine which group is significantly higher (or lower) than the other, then state in words what the differences are.

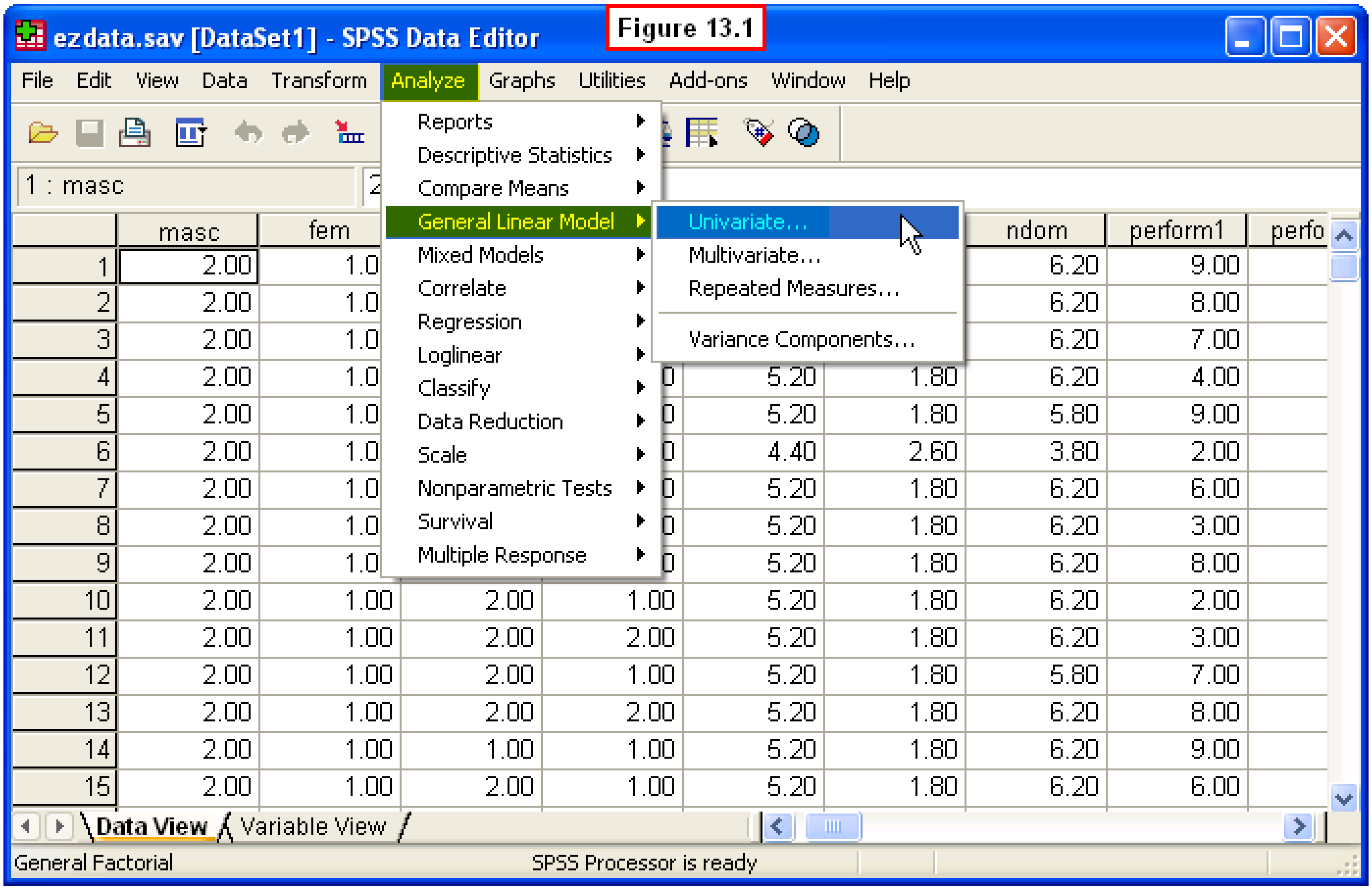
The **interpretation of the interaction effect** is more complex. If the F-value of the interaction effect is *not* significant (i.e., $p > .05$) , then our conclusion would be that gender differences in overall masculinity scores *did not depend* on the level of masculinity (that is, the same gender difference would be seen for both the low-masculinity group and the high-masculinity group).

If the F-value for the interaction *is* significant ($p < .05$), then we would conclude that gender differences in masculinity *depend on* the level of masculinity. For example, it might turn out that when comparing *high-masculine* men to *high-masculine* women, men have higher overall masculinity scores than do women, but that when comparing *low-masculine* men and women, we might see no significant difference in overall masculinity scores between these two groups.

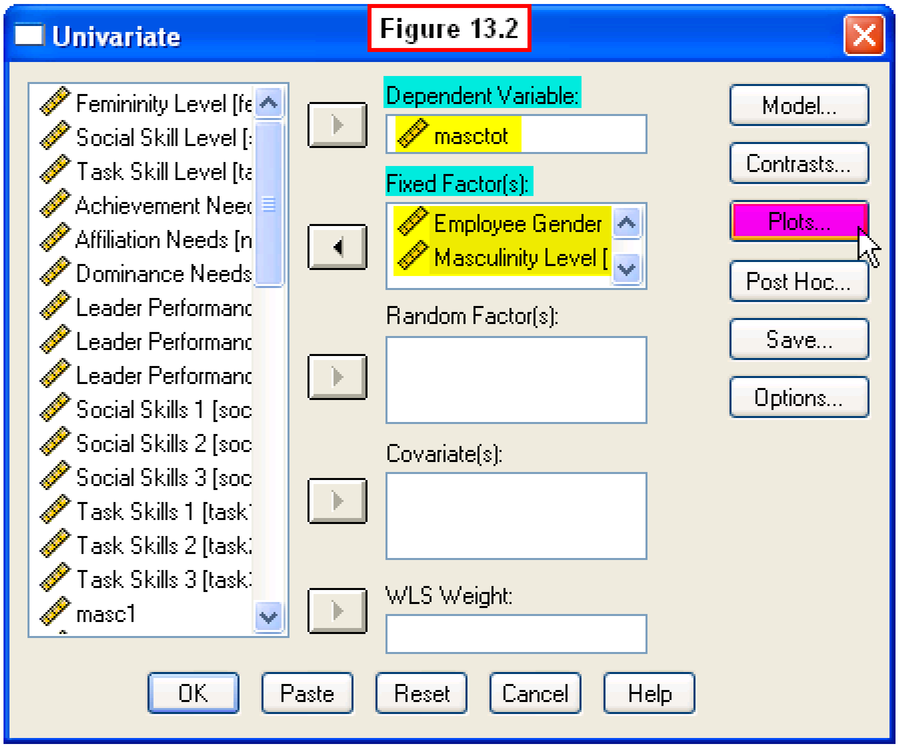
The above example of an interpretation of a significant interaction would require a comparison of the four cell means relevant to the interaction effect (see our discussion of multiple comparisons in Chapters 11 and 12). In our example of the effects of **GENDER** and **MASC** on overall masculinity (**MASCTOT**), this would involve a comparison of the means of each of the four groups described earlier (high-masculine men, low-masculine men, high-masculine women, and low-masculine women). Unfortunately SPSS does not have an option to compute multiple comparisons for factorial ANOVA, so the researcher must do these analyses by hand. For now, suffice it to say that these comparisons would permit us to draw specific conclusions regarding interaction effects, such as those described above.

13.3 Running Two-way ANOVA

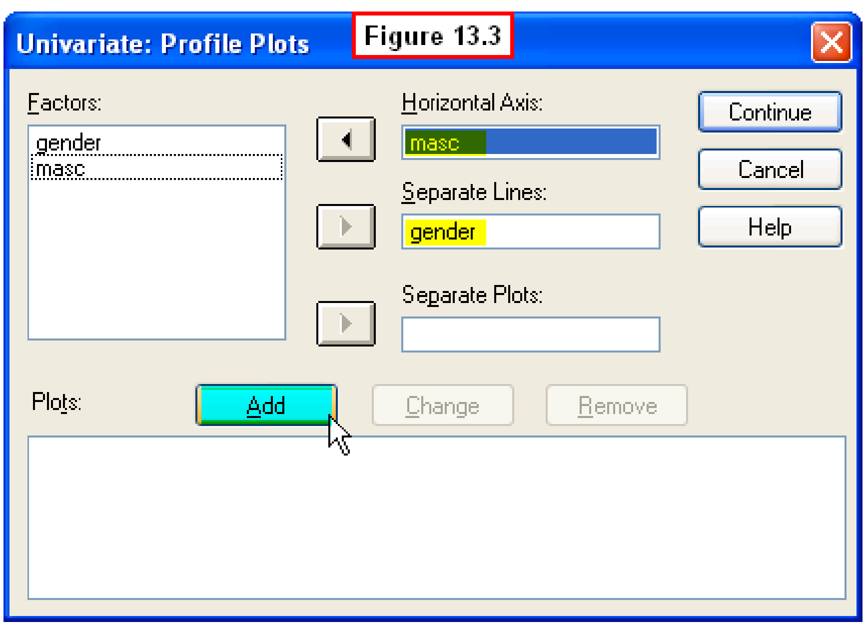
Open your **ezdata.sav** file and select **Analyze, General Linear Model, Univariate...** from the menu (Figure 13.1).



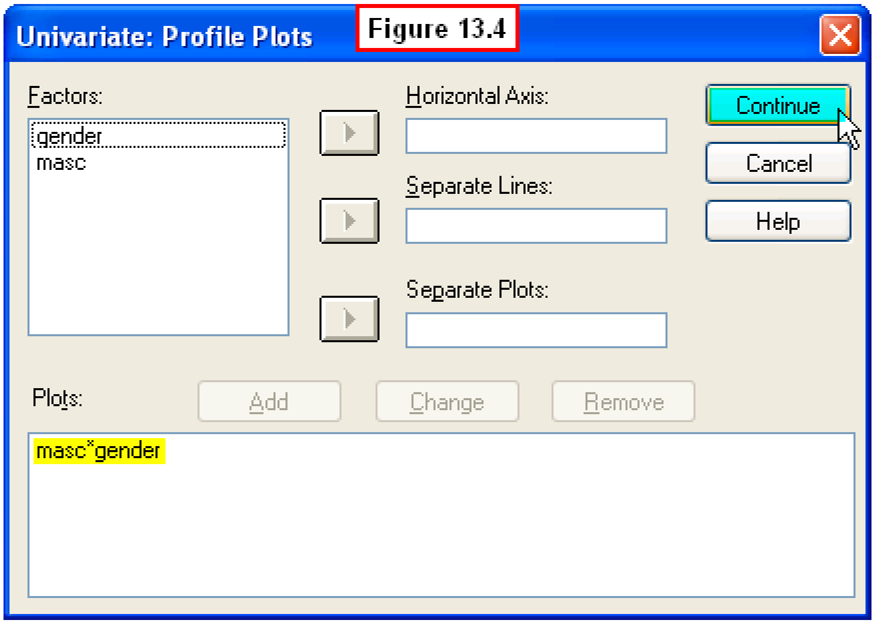
A **Univariate** dialog window will appear (Figure 13.2). In the left pane, scroll all the way down and select **masctot** (overall masculinity score), then move it to the **Dependent Variable:** box on the right. Next, in the pane on the left, select **Employee Gender** and **Masulinity Level (masc)** and move them to the **Fixed Factor(s):** pane on the right (recall that SPSS uses the term, **factor**, to refer to the independent variable, so these terms are synonymous). This will generate the ANOVA table when we run the analysis. In the event that the interaction effect is significant, we will want to generate a figure depicting this to aid our interpretation. To do this, click the **Plots...** button on the right.



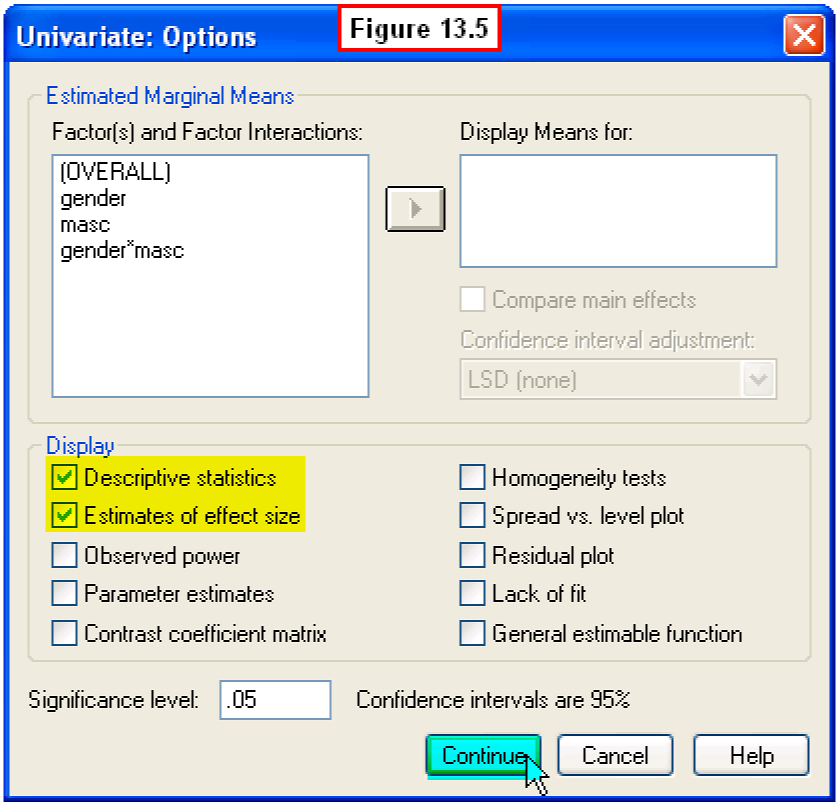
A **Univariate: Profile Plots** window will appear (Figure 13.3). This dialog window wants to know which factor to represent on the horizontal axis and which one to represent by drawing two lines. While technically it doesn't matter which factor is placed where, using separate lines for gender will facilitate comparisons between the sexes. Thus, move **gender** to the **Separate Lines:** box and move **masc** to the **Horizontal Axis:** button.



Next, click the **Add** button at the bottom of this window. You will now see **masc*gender** has been added to the **Plots:** pane at the bottom of this window (Figure 13.4).



Last, click the **Continue** button in the upper right corner. This window will close, returning you to the main **Univariate** dialog window (Figure 13.2). The last thing we need to do is to specify our options. Click the **Options** button on the right of the **Univariate** dialog window. A **Univariate: Options** window will appear (Figure 13.5).



First we need to instruct SPSS to calculate the marginal means for the main effects and cell means for the interaction effect. Note that an estimate of these means could be generated by moving the factors and the interaction from the upper left pane to the **Display Means for:** pane on the right. However, these means will only be accurate if there is an equal sample size in all four cells. This is not the case for our data file.

Further, researchers are also frequently interested in other descriptive statistics (e.g., the standard deviations), and this option only generates means. Thus, instead of using the Display Means option, we will use another way to generate the means (and other descriptive statistics) - the **Descriptive statistics** option. To do this, check the **Descriptive statistics** box in the lower left of this dialog window.

In addition to determining whether or not the independent variables yield significant differences in the dependent variable (which will be assessed by the F-values), researchers typically also want to know how **big** the differences are. To do this, check the **Estimates of effect size** box also. Last, click the **Continue** button at the bottom of this window. This will return you to the main **Univariate** dialog window (Figure 13.2). We are ready to run our ANOVA now, so click the **OK** button in the lower left of the this window.

3.4 Interpreting the Output

The first table in the output (**Between-Subjects Factors**) displays the sample size for each of the two levels of our two factors. As mentioned, the sample sizes are not equal, which is why we chose to generate the relevant means via the **Descriptive Statistics** option. This information is shown in the second table (Figure 13.6). This table presents the marginal means of overall masculinity scores for the two main effects, the cell means for the interaction, as well as the standard deviations and sample sizes. The most relevant information for our purposes are the means.

Figure 13.6

Descriptive Statistics

Dependent Variable: masctot

Employee Gender	Masculinity Level	Mean	Std. Deviation	N
Male	Low-Masculinity	11.8276	2.30014	29
	High-Masculinity	30.0617	.99179	81
	Total	25.2545	8.19872	110
Female	Low-Masculinity	10.1525	2.08288	59
	High-Masculinity	26.0339	.31981	59
	Total	18.0932	8.11139	118
Total	Low-Masculinity	10.7045	2.28518	88
	High-Masculinity	28.3643	2.14312	140
	Total	21.5482	8.89104	228

The overall marginal means for male vs. female employees are highlighted in blue, and the marginal means for the two levels of masculinity category are highlighted in green. Last, the four cell means relevant to the interaction are highlighted in yellow. Before we can interpret these means, however, we must first examine the results of the ANOVA displayed in the **Tests of Between-Subjects Effects** table (Figure 13.7).

Figure 13.7

Tests of Between-Subjects Effects

Dependent Variable: masctot

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	17460.081[a]	3	5820.027	2691.405	.000	.973
Intercept	75511.824	1	75511.824	34919.583	.000	.994
gender	402.875	1	402.875	186.305	.000	.454
masc	14417.369	1	14417.369	6667.148	.000	.967
gender * masc	68.572	1	68.572	31.710	.000	.124
Error	484.389	224	2.162			
Total	123811.000	228				
Corrected Total	17944.469	227				

You can ignore the first two rows of this table and focus on the three rows highlighted in yellow. As we saw in the previous chapter, the most important columns in this table are the last three, which present the F-values (**F**), the significance levels (**Sig.**), and the effect size indices (**Partial Eta Squared**). These present the tests of the two main effects and the interaction effect. Since **Sig.** = .000 for all of these, we can conclude that both main effects and the interaction effect are statistically significant ($p < .05$).

The significant main effects tell us that there are significant differences in overall masculinity score between men and women and between employees in the low-masculine vs. the high-masculine category. The significant interaction effect indicates that the differences between men and women depend on the masculinity category. As discussed in Chapter 12, the partial eta squared values indicate the size of the differences. We can conclude that the size of the difference in overall masculinity between men and women is moderate, and that the size of the difference in overall masculinity between employees in the low-masculine vs. high-masculine categories is huge! This shouldn't be surprising, since the overall masculinity score was used to create the categorical variable, low-high masculinity (see Chapter 5). In fact, the main reason we used the masculinity category variable was to investigate its possible interaction with participant sex.

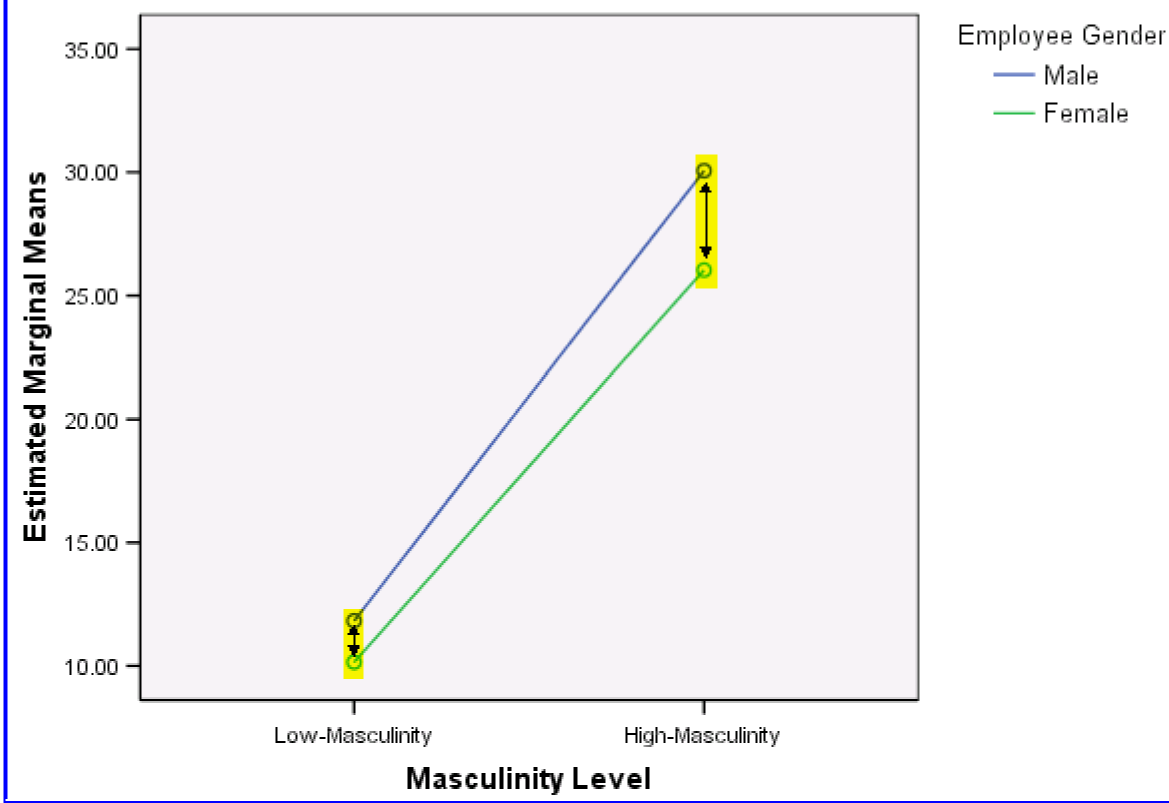
Recall that to provide a complete interpretation of the results, the researcher needs to explicitly state what the differences between the means are (see Figure 13.6). Note that since there are only two levels of our two independent variables, it is not necessary to conduct comparisons between the marginal means, i.e., the F-value already tells us that they are significantly different. However, multiple comparisons would be needed to interpret the differences between the four cell means relevant to the interaction effect (see Chapters 11 and 12). For now, however, below is an example of how these results concerning the main effects would be written in APA format:

A significant main effect was obtained for gender, F (1,224) = 186.31, $p < .001$. Male employees had significantly higher masculinity scores (M = 25.26) than did female employees (M = 18.09). This was a moderate difference (Partial Eta Squared = .45). A significant main effect was also obtained for masculinity level, F (1,224) = 6667.15), indicating that employees in the high-masculinity category had significantly higher masculinity scores (M = 28.36) than did employees in the low-masculine category (M = 10.71). This was an extremely large difference (Partial Eta Squared = .97).

Neither of these effects should be surprising in light of the previous analyses in this text concerning gender differences in masculinity. And, of course, the main effect for masculinity category was guaranteed to occur given that the overall masculinity scores were used to create this independent variable. However, there is more to the story given that the interaction effect was significant, F (1,224) = 31.71, p < .001. Although the effect size is relatively weak (Partial Eta Square = .12), this is actually the most interesting and informative part of this ANOVA, because it indicates that the sex difference in overall masculinity *depends* on the particular masculinity category (low or high).

Interpreting interaction effects and be difficult because it requires the researcher to think of both independent variables at the same time. Due to the complexity of interaction effects, researchers typically graph the four cell means. This visual depiction facilitates comprehension of the interaction, and that is why we generated the graph using the Plots option (see Figure 13.8).

Figure 13.8



The main effect of masculinity level can be seen for both sexes along the X-axis. That is, both men and women in the high-masculinity category scored higher in masculinity than did their counterparts in the low-masculinity category. However, comparisons between the sexes (the blue line for men, the green line for women) shows that the main effect for gender only appears in the high-masculinity category.

Recall that the cell means are highlighted in yellow in Figure 13.6. Assume that we computed multiple comparisons among these four means (unfortunately, SPSS does not provide this option). Here is how the interaction would be interpreted in APA style based on these comparisons:

In the high-masculinity group, male employees had significantly higher masculinity scores (M = 30.06) than did female employees (M = 26.03), reflecting the main effect of gender. However, in the low-masculinity group, men did *not* have higher masculinity scores (M = 11.83) than did women (M = 10.15).

While the implications of this interaction are beyond the scope of this chapter, suffice it to say that two-way factorial ANOVA frequently yields much more interesting information than could be obtained in simpler research designs involving only one factor. Factorial designs allow us to determine qualifying conditions under which simple main effects either do or do not hold. Again, while interpreting interactions can be cognitively taxing, it becomes easier with practice! You will get more practice in the exercise at the end of this chapter!

13.5 Chapter Review Video

Review Me!

13.6 Try It! Exercises

1. Using Two-Way ANOVA: Overall Femininity Differences in Relation to Gender and Femininity Category

Use the procedures described in Section 13.4 to compute a 2 x 2 ANOVA to examine differences in overall femininity (i.e., **FEMTOT**, the dependent variable) in relation to EZ employees' sex (**GENDER**) and femininity category (**FEM**), the two independent variables.

Thus, the four cells of the factorial combination of this **2 (Gender) x 2 (Femininity)** factorial design would result in the following four groups of EZ employees:

(1) Low-feminine males
(2) High-feminine males
(3) Low-feminine females
(4) High-feminine females

After you have completed this two-way ANOVA:

- **Print** your output file to submit to your instructor.
- Answer the following questions for the affiliation variable:
  - Is the main effect of gender significant, and how large is the difference?
  - Is the main effect of femininity category significant, and how large is the difference?
  - Is the GENDER x FEM interaction significant, and what does this mean?
  -
- **Write an interpretation of the results** along with your answers to the above questions to submit to your instructor.
  - Follow the example of APA write-up in Section 13.4 to write your interpretations of the two main effects.

- Assume you conducted multiple comparisons of the four cell means relevant to the interaction effect. Write an interpretation of the interaction graph following the example in Section 13.4.

Chapter 14

Mixed-Model Factorial ANOVA: Combining Independent and Correlated Group Factors

14.1 Introduction to Mixed-Model Factorial ANOVA

In Chapters 9 and 10 we distinguished between two distinct applications of the t-test: the independent samples t-test and the correlated samples t-test. Similarly, in Chapters 11 and 12 we distinguished between independent and correlated samples one-way ANOVA's. Recall that when a **between-subjects design** is used, the appropriate statistical test to use assumes **independent-samples**, whereas **within-subjects designs** require statistical tests assuming **correlated samples**.

In Chapter 13 we introduced two-way factorial ANOVA involving independent samples designs for both independent variables. However, it is possible to have experimental designs involving two independent variables that are both within-subjects. So the same distinctions we made between the two types of t-tests and one-way ANOVA's can be applied to two-way factorial ANOVA.

The researcher must know his/her experimental design in order to run the appropriate statistical test. In this chapter we introduce a two-way ANOVA which combines one independent samples factor and one correlated groups factor. These types of designs are called **mixed-model ANOVA's**, since they involve a mixture of one between-groups factor and one within-subjects factor.

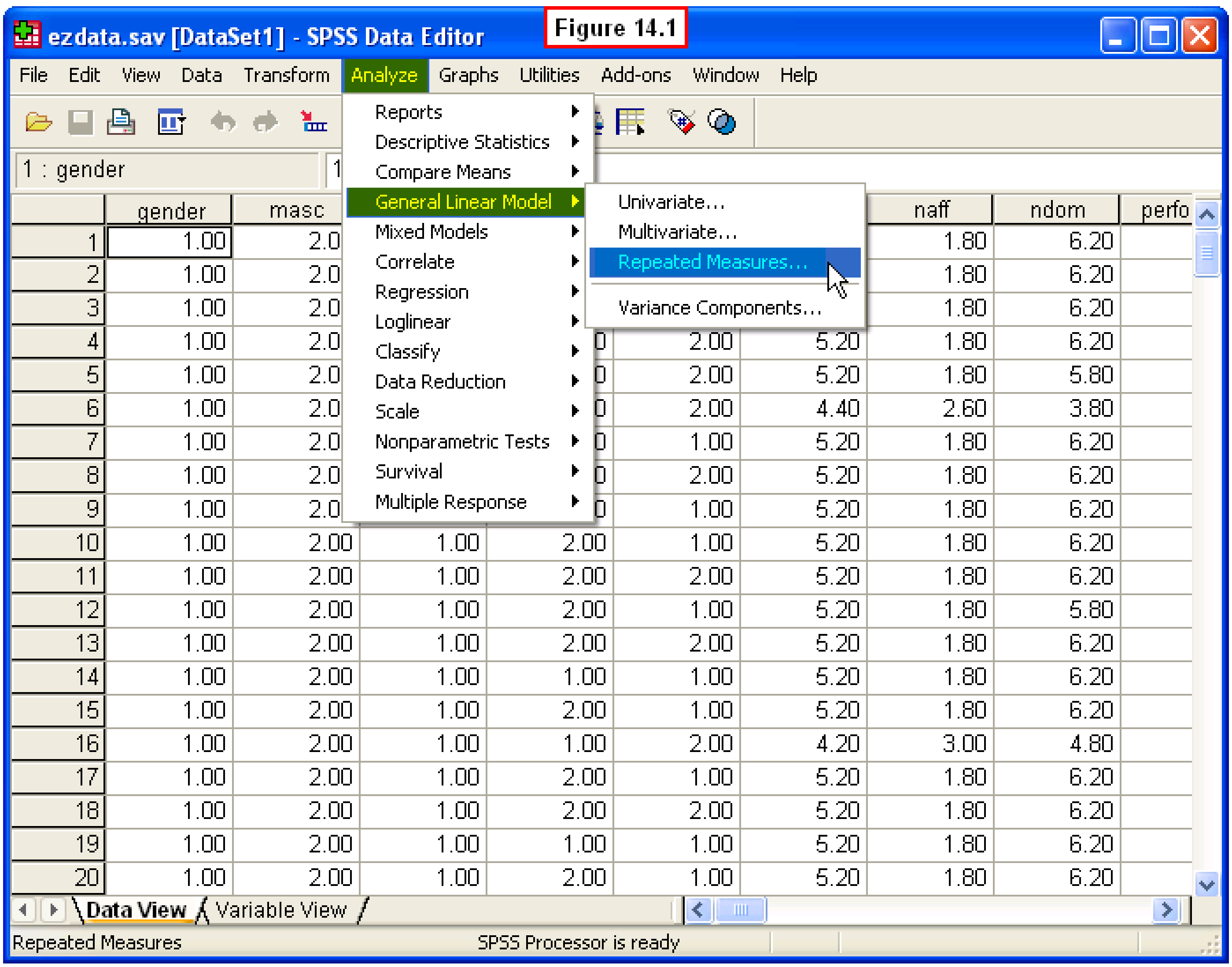
Recall from the past several chapters that we analyzed changes in leader performance, task skills, and social skills before and after attending a leadership training workshop. Since repeated measures were obtained from the sample participants on these variables, they are correlated samples (within-subjects) variables. We saw from these analyses that there were significant improvements from time 1 to time 2 on all of these variables.

Given the focus on gender and gender role stereotypes in this project, an interesting question that could be asked is whether or not the gains on these variables were the same for male employees compared to female employees. For example, it may be that men benefit more from training in social skills than do women, given that women are already relatively strong in this aspect of leadership given their socialization. Further, women may benefit more than men from training in task skills, given the socialization of task skills in men.

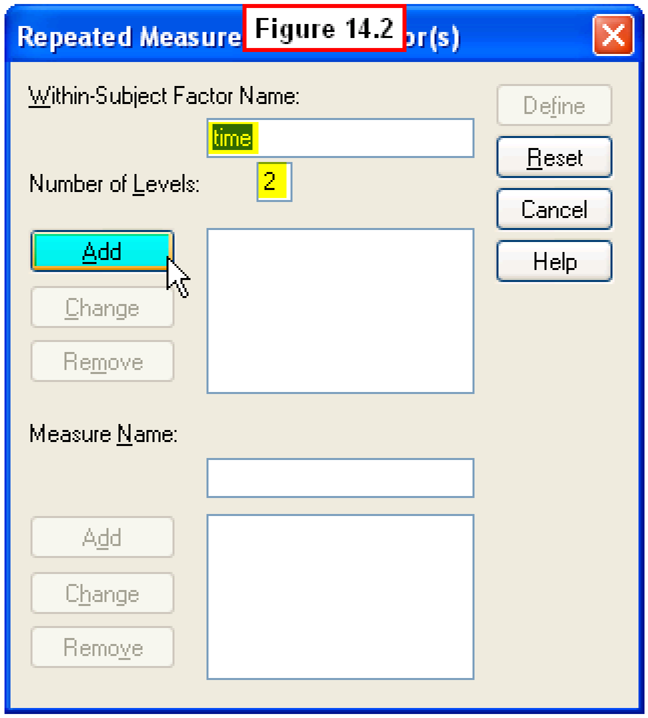
The answer to these questions can be obtained by conducting a mixed-model factorial ANOVA on the social skills and task skills of employees. The example in this chapter will examine the effects of the workshop on task skills in men and women, while the end-of-chapter exercise will ask you to do the same analysis for the effects of the workshop on social skill scores.

14.2 Running Mixed-Model ANOVA

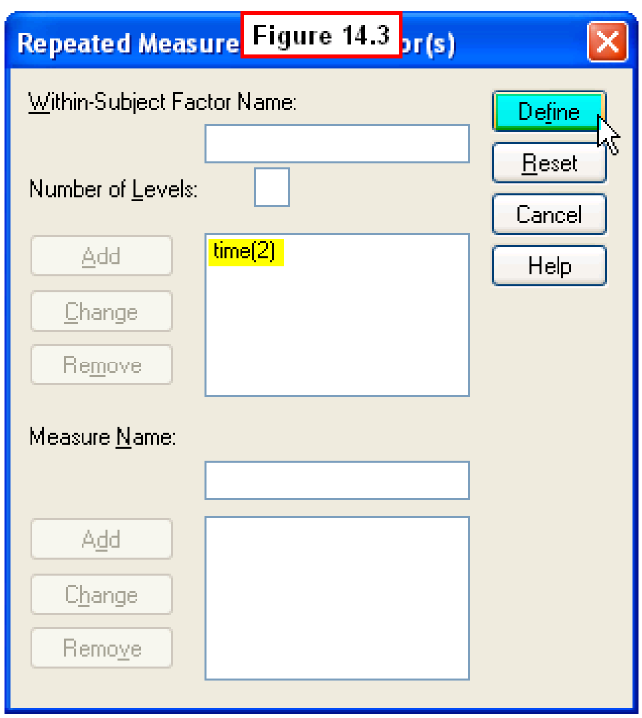
Open your **ezdata.sav** file and select **Analyze, General Linear Model, Repeated Measures...** from the menu (Figure 14.1).



A **Repeated Measures Define Factor(s)** dialog window will appear. As we did in Chapter 12, we need to define the **Within-Subject Factor Name** (i.e., the **Task Skills** repeated measures variable, consisting of the before-after measures, **task1,** and **task2)**, and enter the number of levels of this factor. To do this, type **time** in the box below **Within-Subject Factor Name**, and enter a **2** in the box next to **Number of levels**. Then click the **Add** button.



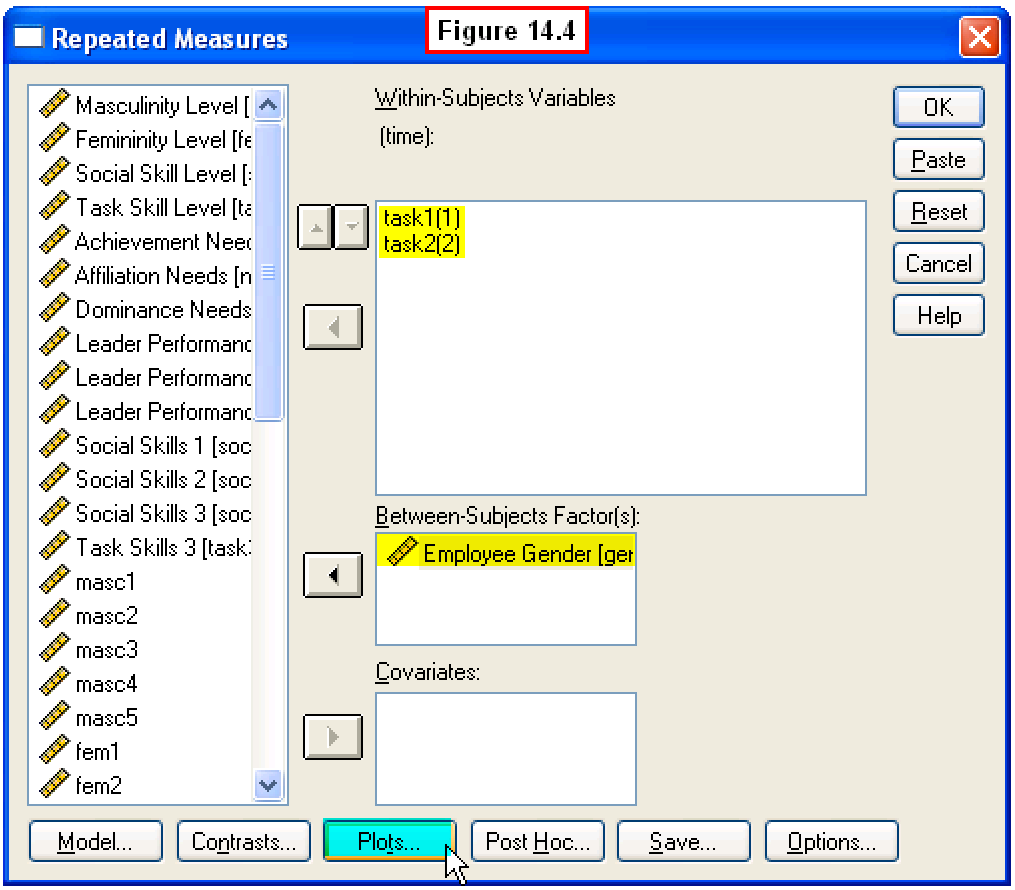
After you have done this, your window should look like that shown in Figure 14.3. Note that SPSS will now name this variable **time**, and it shows that there are three levels: **time (2)**. Now SPSS needs us to define what the two levels of **time** are. To do this, click the **Define** button in the upper right corner.



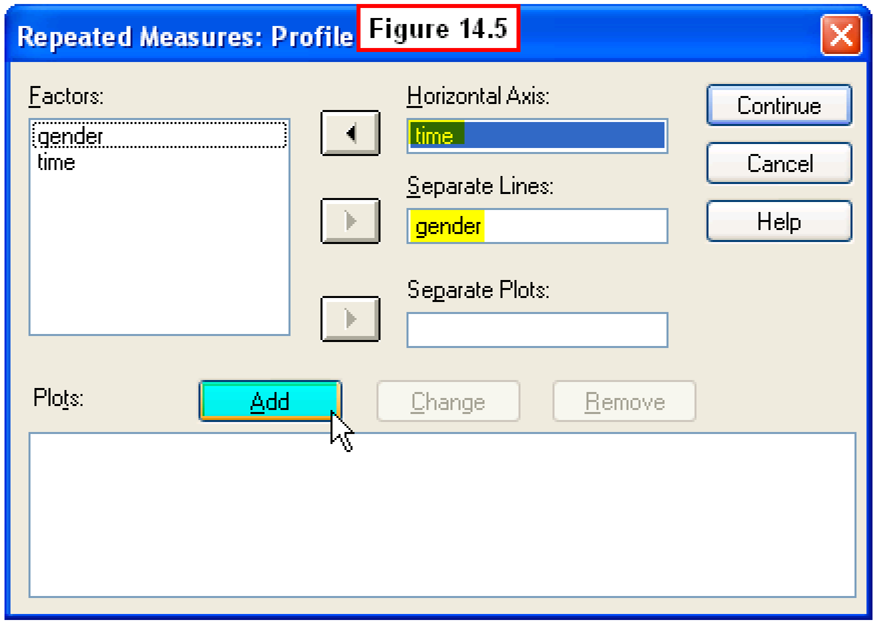
A **Repeated Measures** dialog window will appear (Figure 14.4). To specify the levels, scroll down the variable list on the left, then highlight the first two levels of the **Task Skills** variables (**task1** and**task2**). Click the right-arrow button in the middle, and these variables will be moved to the appropriate places in the **Within-Subjects Variables (time)** variable definition box.

Next we need to specify our **Between-Subjects Factor (Employee Gender)**. Highlight this variable on the left, then click the right-arrow key in the middle to move this to the **Between-Subjects Factor(s):** box on the right.

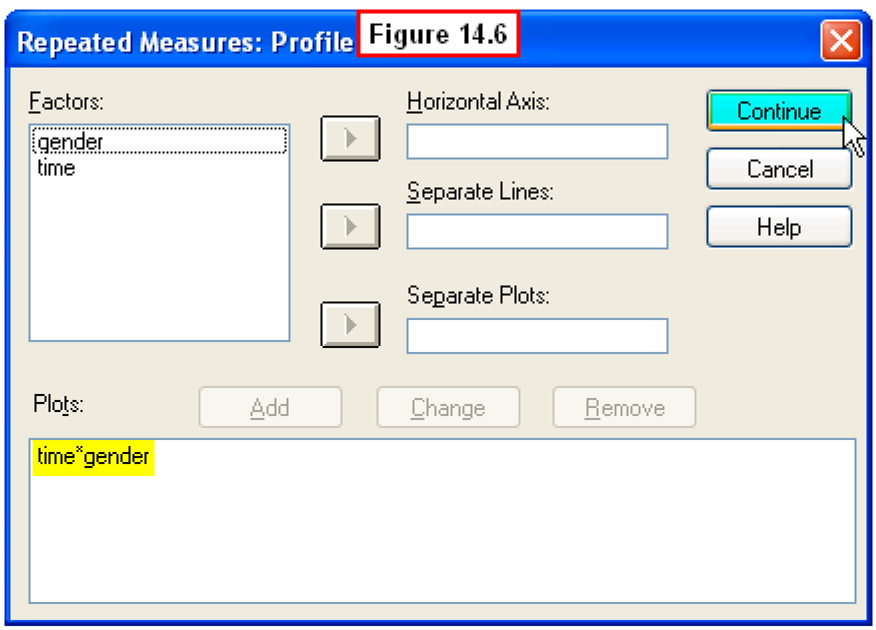
As we saw in Chapter 13, in the event that the **Time x Gender** interaction effect is significant, it will be helpful to generate a graph depicting the interaction. To do this, click the **Plots...** button at the bottom of this window.



A **Repeated Measures: Profile Plots** dialog window will appear (Figure 14.5). Select the **time** variable on the left, then click the right-arrow button in the middle to move this to the **Horizontal Axis** box on the right. Then select **gender** on the left and use the right-arrow key to move it to the **Separate Lines:** box on the right. Last, click the **Add** button at the bottom of this dialog window.

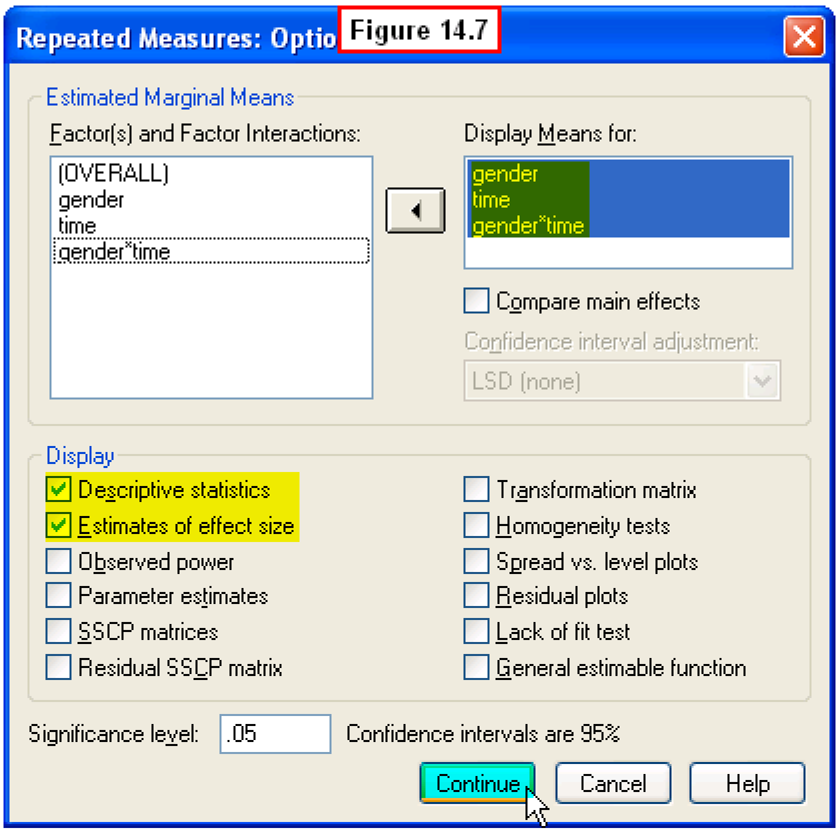


Your window should now look like the one shown in Figure 14.6.



This will cause SPSS to create a figure displaying the mean Task Skills on the Y-axis (vertical) and the two times of measurement on the X-axis (horizontal). SPSS will also generate two separate lines for the mean Task Skills scores before and after the workshop. The first line will graph these two means for men, and the second will depict these means for women. Click the **Continue** button in the upper-right corner, and this will close this window and return you to the main **Repeated Measures** dialog window (Figure 14.4).

Click the **Options...** button in the lower right corner of this dialog window, and a new **Repeated Measures: Options** dialog window will appear (Figure 14.7). To display the marginal means for the two main effects and the cell means for the interaction effect, highlight these in the **Factor(s) and Factor Interactions:** box on the left, and use the right-arrow key in the middle to move them to the **Display Means for:** box on the right (see Chapter 13 for a review of main effects and interaction effects). To generate descriptive statistics and estimates of effect size, check these two boxes in the lower left corner of this window. Last, click the **Continue** button at the bottom of this window.



This window will close and return you to the main **Repeated Measures** dialog window. To run our analysis, click the **OK** button in the upper right corner of this window. The results of this analysis will then be displayed in an output viewer window.

14.3 Interpreting the Output

The first two tables simply list the two levels of the **time** variable and the sample size for male and female employees. Several statistics are presented in the next table, **Descriptives** (Figure 14.8). The most relevant for our purposes are the two marginal means for Task Skills (highlighted in blue) and the four cell means representing the before-after task skills scores of men and women. We will return to a discussion of these means in discussing the interpretation of our ANOVA results.

Figure 14.8

Descriptive Statistics

	Employee Gender	Mean	Std. Deviation	N
Task Skills 1	Male	5.4091	2.01961	110
	Female	4.8559	1.74569	118
	Total	5.1228	1.89895	228
Task Skills 2	Male	5.4545	2.22016	110
	Female	5.5424	1.64677	118
	Total	5.5000	1.94075	228

You can skip over the next two tables of the output (**Multivariate Tests** and **Mauchly's Test of Sphericity**), since they are beyond the scope of this chapter. The next table, **Tests of Within-Subjects Effects**, presents the ANOVA results for the main effect of our within-groups factor, **time**, and the **time x gender** interaction effect (Figure 14.9).

Figure 14.9

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	15.248	1	15.248	7.394	.007	.032
	Greenhouse-Geisser	15.248	1.000	15.248	7.394	.007	.032
	Huynh-Feldt	15.248	1.000	15.248	7.394	.007	.032
	Lower-bound	15.248	1.000	15.248	7.394	.007	.032
time * gender	Sphericity Assumed	11.695	1	11.695	5.671	.018	.024
	Greenhouse-Geisser	11.695	1.000	11.695	5.671	.018	.024
	Huynh-Feldt	11.695	1.000	11.695	5.671	.018	.024
	Lower-bound	11.695	1.000	11.695	5.671	.018	.024
Error(time)	Sphericity Assumed	466.086	226	2.062			
	Greenhouse-Geisser	466.086	226.000	2.062			
	Huynh-Feldt	466.086	226.000	2.062			
	Lower-bound	466.086	226.000	2.062			

The most relevant portions of this table are the F-values, significance levels and effect sizes. The **Sig** column reveals probabilities for both the **time** main effect (.007) and the **time x gender** interaction (.018) are both less than .05, so we can conclude that these are both significant effects (although the Partial Eta Squared values indicate that both are weak effects).

Also skip over the next table on the output (**Tests of Within-Subjects Contrasts**), since it is not relevant to our purposes. The next table, **Tests of Between-Subjects Effects**, presents the ANOVA results for our between-groups variable, **gender** (Figure 14.10). Since the probability in the **Sig.** column (.28) is greater than .05, we can conclude that the main effect for gender is not significant.

Figure 14.10

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	12868.137	1	12868.137	2444.651	.000	.915
gender	6.164	1	6.164	1.171	.280	.005
Error	1189.617	226	5.264			

The next section of the output, **Estimated Marginal Means**, presents information which is partially redundant with the means displayed in Figure 14.8. This section organizes the means into three tables, one for the marginal means of each of the two main effects and a third table which displays the cell means for the interaction effect. The marginal means for the main effect of gender are shown in Figure 14.11.

Figure 14.11

1. Employee Gender

Measure: MEASURE_1

Employee Gender	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Male	5.432	.155	5.127	5.737
Female	5.199	.149	4.905	5.493

Recall that the main effect of gender was not significant ($p > .05$). So even though the mean task skills score for men (5.432) appears to be greater than that for women (5.199), this is not a statistically significant difference. Thus, the appropriate interpretation is that there was not a significant difference in overall task skills between men and women. The marginal means for the main effect of time are shown in Figure 14.12.

Figure 14.12

2. time

Measure: MEASURE_1

time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	5.133	.125	4.887	5.378
2	5.498	.129	5.245	5.752

Recall that the main effect for time was significant ($p < .05$), so it is appropriate to conclude that the mean task skill score after the leadership training workshop was significantly higher (5.498) than the mean task skill score before the workshop (5.133). Thus, as we saw in previous chapters, this analysis also suggests that the workshop was effective in increasing leader task skills.

However, recall that the Time x Gender interaction was also significant. As we saw in Chapter 13, an interaction effect suggests that the effects of one variable depend on the level of the second variable. To interpret the interaction, we need to examine the four cell means (Figure 14.13).

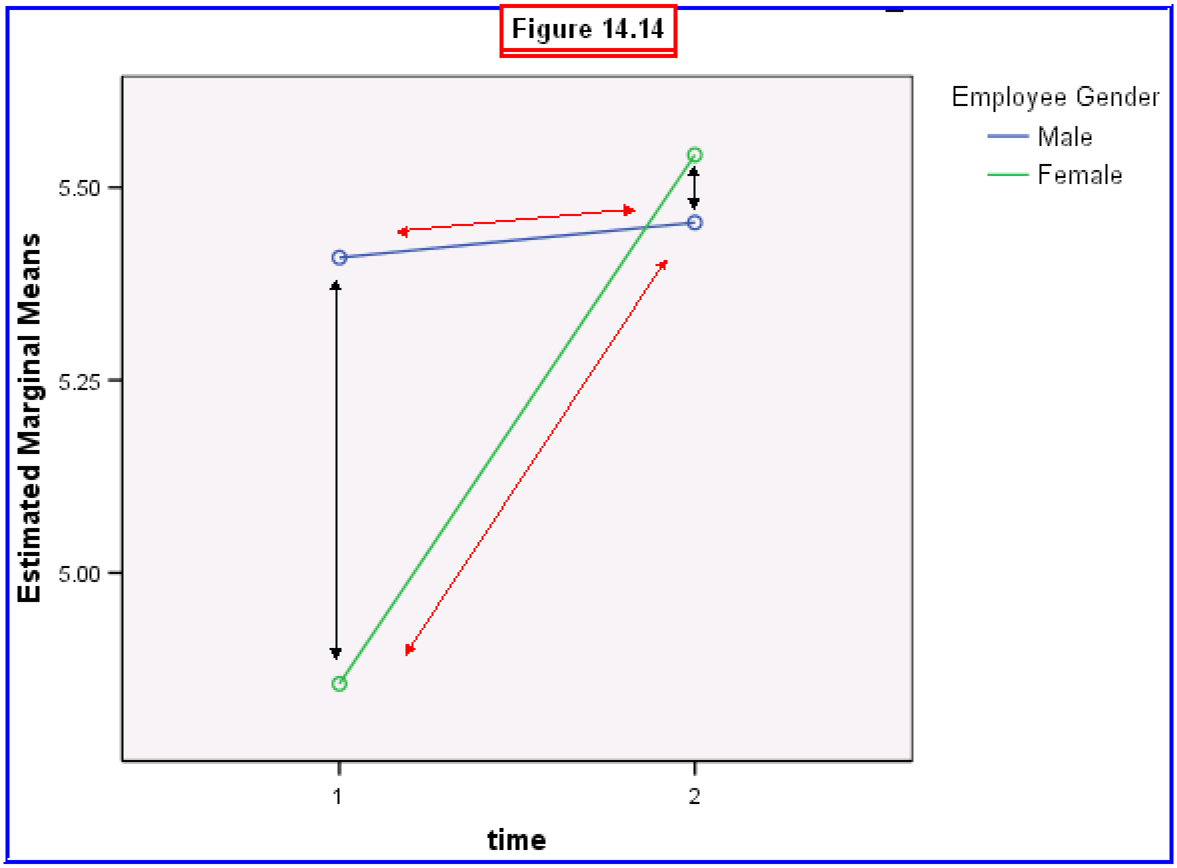
Figure 14.13

3. Employee Gender * time

Measure: MEASURE_1

Employee Gender	time	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Male	1	5.409	.180	5.055	5.763
	2	5.455	.185	5.089	5.820
Female	1	4.856	.173	4.514	5.197
	2	5.542	.179	5.190	5.895

Recall from Chapter 13 that interpretation of an interaction is facilitated by graphing the four cell means. These are depicted in Figure 14.14.



It can be seen in this figure that the effect of the workshop (represented by the Time variable) depended on the gender of the employees. Looking at the two lines, we see that although there was a dramatic increase in task skill scores for female employees from Time 1 to Time 2, the workshop didn't produce much of a change in the task skill scores of male employees.

Further, looking between the lines (i.e., comparing genders at the two times), shows that gender differences in task skills depended on the time of measurement. While men had much higher task scores than did women before the workshop, there wasn't much of a difference between task scores of men and women after the workshop.

In light of gender stereotypes, these results make sense. Men generally are socialized to have high task skills, so they started out at a high level. Thus, the workshop didn't affect their already high level of task skills. Women are socialized to have relatively low task skills, so they started out at a lower level. The workshop was most successful at increasing the task skills of women. In fact, what started out as a big gender difference before the workshop ended up being a trivial difference after the workshop. Below is an illustration of how these results would formally be reported in APA format:

A *2* (Time) x 2 (Gender) mixed-model ANOVA revealed that the main effect for Gender was not significant F (1,226) = 1.17, p > .05, Eta-squared = .01. Thus, there was no overall difference in the task scores of men (M = 5.43) compared to women (M = 5.20).

A significant main effect for Time was obtained, *F* (1,226) = 7.39, *p* < .05, though this was a weak effect (Eta-squared = .03). Task scores after the workshop (*M* = 5.50) were significantly higher than before the workshop (*M* = 5.50).

However, a significant Time x Gender was also obtained, *F* (1,226) = 5.67, *p* < .05, though this was a weak effect (Eta-squared = .02). Examination of the cell means indicated that that although there was a large increase in task skill scores for female employees from Time 1 (*M*= 4.86) to Time 2 (*M* = 5.54), the workshop didn't produce much of a change in the task skill scores of male employees from Time 1 (*M* = 5.41) to Time 2 (*M* = 5.46). Before the workshop men had much higher task scores (*M* = 5.41) than did women (*M* = 4.86). But after the workshop, there was not much of a difference in task scores between men (*M* = 5.46) and women (*M* = 5.54).

Technically, we would have had to compute multiple comparisons among the four cell means to justify the above interpretation of the interaction, but that is beyond the scope of this chapter. Our main goal was to provide you with another example of how to interpret an interaction, rather than to also explain the multiple comparison procedures (which would support these conclusions anyway).

As we saw in Chapter 13, factorial ANOVA can yield much more useful information than simpler types of experimental designs. Based on the present analysis, you could report to EZ execs that the leadership training program benefited female employees much more than male employees, resulting in task skill scores equivalent to those of men after the workshop. Before you recommend only sending female employees to the leadership training workshop, let's see what the effects of the workshop are for social skills scores between men and women. That will be the exercise at the end of this chapter.

14.4 Chapter Review Video

[Review Me!]

14.5 Try It! Exercises

1. Using Two-Way Mixed-Model ANOVA: Effects of Leader Training Workshop on Social Skills Scores of Men and Women

Use the procedures described in Section 14.2 to compute a 2 (Time) x 2 (Gender) mixed-model ANOVA to examine differences in social skills before the workshop (i.e., **soc1**) and after the workshop (i.e., **soc2**) in relation to EZ employees' sex (**GENDER**).

After you have completed this two-way ANOVA:

- **Print** your output file to submit to your instructor.
- Answer the following questions:
    - Is the main effect of time significant, and how large is the difference?
    - Is the main effect of gender significant, and how large is the difference?
    - Is the Time x Gender interaction significant, and what does this mean?
    -
- **Write an interpretation of the results** along with your answers to the above questions to submit to your instructor.
    - Follow the example of APA write-up in Section 14.3 to write your interpretations of the two main effects.
    - Assume you conducted multiple comparisons of the four cell means relevant to the interaction effect. Write an interpretation of the interaction graph following the example in Section 14.3.

Chapter 15

Epilogue:
SPSS, Data Analysis and the EZ Research Project

15.1 Computer-Based Analyses of Research Data

It seems appropriate to conclude with a review of what we have learned in this book about the use of the computer – and the SPSS statistical package in particular – in the analysis of data obtained in a research project such as the hypothetical study at EZ Manufacturing. This text has introduced you to an extremely powerful tool and has helped you develop the skills needed to use it in analyzing data.

In the first several chapters, you learned the basics for using the Syntax Code method and the Point-and-Click method in using SPSS for Windows. You learned how to create a data file consisting of scores from on several distinct variables related to a research question. You learned how to edit that file and to obtain a printout of the file. You also learned how to generate simple frequency distributions of the scores on each variable in the file, as well as how to transform existing variables into newly-created variables. We saw that this procedure could be helpful in getting an overall feel for the data, but that drawing meaningful conclusions from this cursory description of the data was next to impossible.

We then proceeded to the statistical analysis of the data so that we could obtain a deeper understanding of the results. We introduced you to the wide range of descriptive and inferential statistics that could be computed using SPSS on our data file. In Chapter 6, you learned how to generate a variety of **descriptive statistics**, such as the mean, median, mode, range, and standard deviation. This number crunching SPSS routine saves the researcher hours of tedious labor that would be required to compute these statistics by hand. It's also much more accurate. These statistics can be very useful in helping the researcher get to know the data by obtaining single numbers that summarize the entire set of scores for each variable. They are also essential in making comparisons between important groups in the study (e.g., comparing mean leadership performance scores of men versus women).

Chapter 7 introduced you to a routine for creating frequency matrices and contingency tables for scores in various combinations of levels of two or more nominal variables (e.g., the number of task oriented male versus female employees). In Chapter 8 you learned various correlational subroutines that are useful in describing direction and degree of relationships between two quantitative variables (e.g., the correlation between Achievement Needs and Task Skills), and in predicting scores on one variable from another variable.

The above statistical analyses took us a long way toward understanding our data and the relationships between variables relevant to our basic research questions. However, in order to determine whether there were real differences among the various groups being studied, we had to rely on several **inferential statistics** that permit such conclusions. In Chapters 9 and 10 we learned how to use the *t*-test procedure to analyze differences between mean scores of groups on the same dependent variable (e.g., Social Skills scores at two levels of a single independent variable (e.g., Gender).

In Chapters 11 and 12, a more sophisticated procedure was introduced (One-way ANOVA) which allowed the researcher to test for significant differences between three or more groups (e.g., mean Leadership Performance scores of masculine sex typed, feminine sex typed, and androgynous employees). We concluded in Chapters 13 and 14 with an extremely powerful procedure (Factorial ANOVA) which permits simultaneous testing for differences between levels of two independent variables (e.g., Gender and Leadership Style) in addition to assessing their interactive, or combined, effects on the dependent variable.

We assume this introduction has given you an appreciation for the value of the computer in helping the investigator answer both simple and complex research questions. The computer not only saves time and effort, but it can enable the researcher to pose and answer questions that would be difficult, if not impossible, to address by eyeballing the data or by hand computations. We believe that the basic skills you have learned in conducting the above analyses on our hypothetical research project will be easily transferable to future projects that you may undertake. Indeed, you should have acquired a better understanding of statistics and research methods from the examples and exercises, and we hope that you will be motivated to apply this knowledge and skill to whatever career you pursue.

This, of course, brings us back to one of the motivating factors for the EZ project in the first place – the mega consulting fee that upper management at EZ Manufacturing is paying you for conducting this study! What are you going to be able to report to them that might be helpful in the development of their affirmative action program? Let's consider some of the conclusions and recommendations that you might make on the basis of your research and data analysis.

15.2 Sex Roles, Leadership Style, and Leader Effectiveness: A Review

Recall from Chapters 4 and 5 that upper management is interested in identifying criteria for making sound decisions about which individuals (especially female employees) to promote to leadership positions. They hired you to conduct a study investigating important variables related to both gender and leadership effectiveness. Your literature review revealed that both sex role stereotypes and leadership style have been related to performance effectiveness in the past, so you measured all these variables in a selected group of men and women at EZ Manufacturing. Further, you obtained measures on other relevant variables, such as leader social/task skills and work motivation needs. After collecting and analyzing the data in the manner described in the preceding chapters, you must now prepare a report to be presented at the next board meeting.

What do you tell the eager execs that will convince them that you are worth your huge fee? Before you panic, recognize that it will not be possible to report everything you have learned from your study to these busy people. It will not be necessary to present all the minute details of methodology and statistics discussed in this book – the execs are going to assume that your conclusions are warranted by sound methods and analyses. Indeed, this is another argument for the value of the computer: it has done all the work that permits you to present a meaningful summary and make recommendations that the executives can understand and implement.

To be sure, some tables and figures will lend credibility to your conclusions and facilitate your presentation. However, we will not concern ourselves here with this issue, nor will we pretend that what follows is a prototype of such a report. Instead, we will simply review some of the important and interesting findings of this hypothetical study as a way of helping you to integrate the diverse analyses that you conducted in the previous chapters. Before going on, it might be useful for you to glance over your printouts and jot down some of your own conclusions to compare with ours.

First, we learned that sex role stereotyping does exist at EZ Manufacturing. Many male employees view themselves as primarily masculine, and many female employees view themselves as stereotypically feminine. However, you were able to identify a group of male and female employees who are androgynous. They see themselves as possessing attributes that are *both* masculine and feminine. You also discovered that there are consistent differences in leadership style among employees: some are primarily task-oriented, others are relations-oriented, and still others combine task and social skills in their leadership style.

Further, you uncovered the interesting fact that most of the masculine sex typed employees are task oriented, most of the feminine sex typed employees are relations oriented, and most of the androgynous employees combine task and relations orientations in their leadership style. Thus, traditional stereotypes of a task-oriented leadership style in men and a relations-oriented style in women appear to hold primarily for sex typed employees.

While the above results are interesting in themselves (as well as theoretically important from a scientific point of view), the EZ executives are now anxiously awaiting your interpretations and conclusions regarding their planned affirmative action program. The results concerning your measure of leadership performance effectiveness are key here. Indeed, you discovered many worthwhile relationships to report. One important finding is that there were no significant overall differences between men and women in performance effectiveness. Thus, there is no basis for anxiety about promoting women to traditionally male leadership positions. Thus, stereotypes about men and leadership do not appear to hold at EZ Manufacturing; women appear to be equally as capable as men, and you can recommend that the affirmative action program be implemented without hesitation.

But which women (and men, for that matter) should be promoted? Your results demonstrated that androgynous men and women are more effective than sex typed employees, so this is a good group to target for promotion. Further, you have evidence as to why this relationship holds. Recall that androgynous employees were more likely to exhibit a combination of task and relations strategies than were sex typed individuals. Other analyses revealed that individuals exhibiting the combined leadership style are more effective than are people who are exclusively task or relations oriented.

Thus, the overall pattern of results suggests that the affirmative action program (and promotions for male employees as well) is most likely to be successful by promoting those individuals who are androgynous and/or exhibit a combination leadership style. It is important to emphasize the value of this knowledge: the net effect of this strategy will be to increase the effectiveness, productivity, and profit potential of EZ Manufacturing as a whole by having the best possible people in leadership positions (a conclusion that will be music to the ears of EZ execs). Finally, you might conclude your report by noting some of the interesting results of other variables, such as work motivation needs, that might assist in developing recruitment, hiring, and promotion guidelines within the corporation.

15.3 Caveat and Conclusion

We believe that enough has been said to give you a feel for the kinds of conclusions and recommendations that could result from a study such as the hypothetical one presented in this book. As we have emphasized repeatedly, these data are hypothetical, and your authors can now confess that we constructed them so that they would come out this way.

While the results have not been concocted out of the blue, the real world of research is not so likely to be as neat and clean. The data do correspond with some actual research in the literature, but you should note that there is never a guarantee that this will be the case – many major projects dead end with inconclusive results, a possibility that you should be prepared for and certainly should warn your clients about before agreeing to conduct the study!

Our goal was to construct a data set that would be manageable for leading you through the various procedures introduced in this text and would yield meaningful results that would be fairly easy for you to understand. We hope that we have accomplished this goal. We also hope that in the process you have acquired the skills needed to address research questions like the ones presented by our EZ project, or to obtain empirically-based answers to any other questions which might stem from your own particular interests. The tools of the computer and SPSS and the skills for using them have been introduced here – it remains for you to take advantage of them!

Appendix
Reassurance & Help For the Nervous Novice

David Rowland
Valparaiso University

1.1 Goals of this Appendix

Computer technology has permeated every phase of research and experimentation in the social and health sciences; and within related fields such as business and education, the computer has become an indispensable tool, from marketing research and consumer behavior to trend analysis and forecasting.

One of the first and primary uses of the computer for social scientists has been that of ***data analysis***. Clearly, it is this single aspect of computing that has enabled social scientists to grasp the truly complex nature of human behavior and social organizations. It is not surprising, then, that as students of statistics, our focus is on the computer's ability to manage, process, and analyze data.

Nevertheless, social and health scientists have come to involve the computer in nearly every component of their research, from the literature search, to the implementation of experiments, and to the writing of the manuscript on a word processor.

There are several elementary goals in this appendix: we first review several basic ideas about the nature of research and different kinds of research. Then we discuss the critical role that computers play in this process. We give you just enough understanding of computer functioning so you can feel computer-literate. Next, we provide a general introduction to statistical software packages used for data analysis. Finally, we explain some of the terminology we use in providing instructions throughout the text and illustrate ways you can access help from within SPSS.

1.2 A Word of Reassurance

For most recent graduates of high school, computers are familiar friends. But for those who seldom use them, computers can be intimidating. The beginner is sometimes fearful of breaking the computer, making errors, ruining or losing important files, or just being confused. Whether or not you're a beginner, as you learn any new application of the computer, you will encounter your share of frustrating situations. In this book, for example, you will learn to use a statistical software package (we'll define that later), and initially, you may find all the steps necessary for creating, modifying, saving, and running files somewhat overwhelming.

In the long run, though, it's easier to learn about these procedures with the guidance of an instructor, classmates, and an instruction text than it is to learn them on your own. To be successful, you will need instruction (provided by your instructor and this book), computer facilities (provided by your institution), and determination (that's your contribution).

But be assured, the skills that you are about to learn here will serve as a valuable asset in many professional jobs and will be essential to successful graduate study in nearly all fields of health and science. If you're a bit apprehensive, you're normal. Just realize that as you begin to use the computer, you will not understand everything at once. So you need to follow two simple rules:

- When you are confused, ask a question. Most questions about computers have relatively simple answers, and you can save a lot of time by asking for help rather than by trying to solve problems on your own. Your instructor is the most obvious source of help, but many of your classmates who have had computing experience will also be glad to help (in fact, you'll find that some students like to demonstrate how much they know!). Your questions will most likely be those of a typical beginner, so don't worry about sounding "stupid."
- Take the time to jot down the answer because otherwise you will soon forget it.

1.3 Bivariate and Multivariate Correlational Research

In any field of science, research represents the way in which predictions are tested, theories are developed, and the knowledge base is expanded. Without research, science would not exist, and so understanding research methodology is the key to understanding how scientists come to know what they do. Only by understanding the research process can you grasp the power of the scientific method while also realizing the limitations (and possibility of errors) inherent in the process.

Scientists use a variety of methodologies in their pursuit of knowledge, and the social and behavioral sciences employ just about all of them. Some social science research is aimed primarily at description - the researcher may be interested in describing people's responses on a number of individual variables. For example, a political scientist may want to characterize attitudes toward the President and his social policies. Or a sociologist may want to describe the characteristics of a particular religious cult. In either case, the researcher would choose a variety of statistical measures, called univariate and multivariate procedures, to describe responses on each of these individual variables.

Typically these statistical measures would include indices of centrality (e.g., a mean or median) and dispersion (e.g., the range or standard deviation). Conceptually, these are not complex statistics; but computation - actually calculating them - may require tedious (and error-prone) work. Who would want to calculate by hand (or even with a calculator) the mean response to four questions taken from 2,200 people, a sample size not uncommon in some areas of social science research?

Beyond simple description of individual variables, a great deal of research attempts to show that two variables are in some way related to one other. Often, this type of research is also descriptive in nature, but rather than describing single variables, it describes relationships between two or more variables. If the procedure attempts to relate just two variables at a time, we refer to the techniques as **bivariate**. For example, a sociologist might attempt to show that there is a relationship between one's attitudes about a President and one's income level. Or a nutritionist may demonstrate a correlation between one's daily sodium intake and blood pressure. In this type of research, the investigator takes measures on two variables and determines how strongly they are linked and, in some cases, whether they are positively or negatively related.

Common bivariate strategies with which you may be familiar include crosstabulation and correlation. As you may have already learned, although a researcher may be able to demonstrate a strong relationship between variables with correlational methodology, s/he is never justified in concluding causality between variables. Monthly sales of winter coats correlates negatively with the number of building construction injuries, yet obviously there is no causal connection between these two variables. Can you identify the mysterious unknown variable (sometimes called a lurking variable!) that explains the changes in both variables? (Hint: everybody talks about it, but no one ever does anything about it!)

Of course, the correlational approach to understanding behaviors, attitudes, or responses need not be limited to the investigation of just two variables at a time. One's blood pressure (or other measures of cardiovascular health) may be related to many variables besides sodium intake, including exercise, water intake, other dietary factors, family history, medications, and so on. Statistical procedures which determine how several variables might relate to one or more variables of interest (e.g., blood pressure, and perhaps other indices of cardiovascular health) are referred to as **multivariate** procedures. Such procedures can be conceptually quite complex, and computationally, they are nearly impossible to perform without a computer.

1.4 Experimental Research

A third research strategy, called the experimental method, represents the scientist's way of trying to establish a cause-effect relationship between two or more variables. In a classic experimental design, the researcher systematically manipulates or changes the variable presumed to be the *causal agent* (this variable is called the **independent variable**) and then determines whether there are corresponding changes in the variable under observation (the *presumed effect* called the **dependent variable**).

Consider the previous correlational example investigating the relationship between sodium intake and blood pressure. If we should in fact find a correlation between these variables, we might then be interested in trying to establish a causal link between them. The experimenter might randomly assign participants to groups that receive different amounts of daily sodium intake (this represents a way of manipulating or changing the value of the independent variable) and compare blood pressure (the dependent variable) across the groups after a specified period of time.

The experimenter would also attempt to control unwanted or extraneous variables (e.g., family history of cardiovascular disease) which might impact the dependent variable by holding these variables constant (e.g., equating the groups on family history). These are sometimes called **extraneous variables**, or if held constant, they may be called **control variables**. By using the experimental method involving manipulating and independent variable while controlling extraneous variables, the researcher could indeed establish a stronger causal connection between the two variables of interest, in this case, sodium intake and blood pressure. Examples of statistics used in experimental research include the *t*-test and Analysis of Variance (ANOVA).

Just as with correlational procedures, experimental methods need not be restricted to the investigation of two variables at a time. They too may be multivariate, in that the effects of several independent variables may be investigated on one or more dependent variables simultaneously. As you might expect, these procedures can become quite complex, and are invariably simplified with the aid of a computer.

The aforementioned distinctions among descriptive, correlational, and experimental methodologies are convenient categorizations for the beginning student, but in fact these clean distinctions are often lost in actual research situations. Some experimental designs are called quasi experimental because they use independent variables which cannot be manipulated (e.g., race, eye color, or IQ).

These designs are really more correlational than experimental in that they do not permit strong cause-effect conclusions. Some multivariate procedures combine correlational and experimental strategies (an example is a procedure called analysis of covariance). And indeed, some statistical experimental designs may be considered a variation of the "general linear model," one which is most easily conceptualized in terms of multiple regression, a correlational technique.

While this overlap in methodologies is certainly confusing to the beginning student of statistics, one point clearly emerges. Research strategies and ensuing statistical procedures come in all shapes, sizes, and variations. These more subtle distinctions, however, are not critical for the beginning student of science; nevertheless, when dealing with multivariate procedures, their understanding becomes more useful, and so we will reiterate some of these distinctions when we approach those topics.

1.5 So Where Do Statistics and Computers Fit In?

No matter what scientific methodology is used, the variables under investigation must always be measured in some way or another. And measurement generates numbers - or data - often much more data than we can readily interpret by quickly scanning the individual scores. In multivariate psychological or sociological studies, as many as 50-500 variables on over several hundred participants may be measured in a single study.

As a result, statistical tools have been devised to help the researcher summarize and interpret raw data. Statistical procedures not only provide a way of summarizing data into quick and manageable information (if you're in a hurry, knowing the mean IQ of 100 ninth graders is probably more meaningful than viewing all 100 IQ scores), but they also help the researcher make decisions about whether relationships between two or more variables, such as in our example with blood pressure and sodium intake, are actually real ones.

Just as statistical methods provide us with a tool for transforming data into manageable information, computers enable us to use statistical tools of greater complexity and with greater effectiveness. One can crush ice cubes, for example, by beating them with a rolling pin, or by using the blender--which would you prefer? Computers offer the same type of advantage when one is dealing with the processing of information. Why use a calculator, or rely solely on your own mental capabilities, when computers can achieve the same ends with a lot less agony? Many jobs that require several hours of computer time would require days, even months, of human time. In fact, some tasks, because of their complexity and length, simply could not be performed without the help of the computer.

But another advantage is that computers do not make mistakes. So called computer errors are usually nothing more than human errors (e.g., giving the wrong command, a mistake in the software program, etc.) that are attributed to the computer. In nearly all cases, the computer is merely doing what it has been instructed to do, rather than what it should have been instructed to do. Finally, the computer never tires or complains about tedious or repetitive tasks. Have you ever had to balance several months of check writing in your checkbook? As you know, even these relatively simple tasks can end up being rather arduous, especially if your balance and the bank's balance don't match. Just imagine how tedious the bank's job would be if its staff had to perform all the calculations for the checking accounts by hand, or with calculators!

So, while computers may be sophisticated pieces of machinery beyond your immediate comprehension (you don't really have to know what a chip is or how many bytes are on the hard disk in order to use a computer effectively), they function simply as tools that can help you do your job more easily, efficiently, and effectively. As with any tool, you, the apprentice, may need to understand some general characteristics of the tool in order to use it effectively, and, as you might expect, the more training you have, the easier it will be to master its use.

Now that you know that the computer is a tool, what does this tool work on? Is a computer really analogous to, say, a more conventional tool such as a table saw or blender? In a manner of speaking, yes. Just as these conventional tools can help transform raw material into finished products (wood into a desk, or vegetables into a puree), a computer starts with the raw material of data and produces a transformed product in the form of information. Although computers can also perform a wide range of other functions, the transformation of numbers (data) into information represents the traditional role of computers. This is exactly what you will be doing with the computer when you learn to use a statistical software package.

Consider the simple example of isolated numbers in Table 1.1, Column A. Here we have **raw data**, a series of numbers with no immediate meaning.

Table 1.1	
A **Score**	**B** **Week's High Temperatures**
71	Monday: 71 degrees
73	Tuesday: 73 degrees
68	Wednesday: 68 degrees
61	Thursday: 61 degrees
65	Friday: 65 degrees
	Week's Average: 67.6 degrees

But now consider the numbers under Column B. Now that headings have been added, the numbers organized, and the average computed, the numbers take on meaning - they have become **information**. In this case, data were processed (though not necessarily with a computer), with the result being information. Essentially, the computer performs a similar function: it takes a series of symbols and processes them so as to allow a meaningful interpretation.

1.6 Answers to Some Basic Questions about Computers

1.6a How do you make the computer do what you want?

Now that you know some basics about computers, let's talk about how we make the computer process data to produce information. Obviously, we need a way to communicate with the computer to make it do what we want. This basic process is accomplished with the use of software. Software is a program or collection of programs used to perform a function. A program is a set of specific instructions for the computer to execute.

A software package refers to the software program, along with other materials such as manuals. However you decide to use the computer, it is the software that enables you to do so. If you want to use the computer to keep track of various business accounts, you need software that allows you to do that. If you want to use the computer for word processing, or data analysis, or producing graphs, you need software that can be used for each different application. Our primary interest is in learning to use a software program for data analysis, but as part of that process, it is often necessary to learn to use other software systems (e.g., Windows) on your computer as well.

With the advent of the first statistical package for data analysis in the mid 1960's, the task of telling the computer what to do and how to do it became greatly simplified for the person interested in data processing.

1.6b What is a statistical software package?

A statistical software package consists of a series of pre-written and pre-tested programs that perform a number of specified operations on data. For example, the software package may have a program that calculates the mean and median for a set of data. Many statistical software packages are currently available.

Since these packages are merely large software programs, they are purchased separately from the computer and separately from each other. They are then stored on disk as part of the secondary or auxiliary memory of the main computer or server. Or if you're using a single-user (PC) system, they may be stored directly on the hard drive. The cost or annual subscription to these packages may range anywhere from several hundred to thousands of dollars.

Each statistical package has its own set of unique capabilities and commands. However, there are common elements to the logic behind almost all statistical packages, so as you learn one system, you'll also find it easier to work with other systems.

1.6 c Overview and characteristics of statistical software packages

The number and types of statistical software packages that are available continue to grow each year. Often these packages are designed to target a particular audience, for example, researchers specifically in the social science researchers, or in biomedical fields, or public health, and so on. Statistical software packages share many common features. For example, most can transform data, create and merge data files, construct new variables, and perform various statistical computations such as means, correlations, t-tests and so on. But not all packages can perform all operations as easily or efficiently, so there may be times when the researcher finds it necessary to use a package with which s/he is not familiar.

A situation like this may arise when the user wants to perform a statistical analysis that is infrequently used by social scientists. Whereas one package may not be capable of performing this analysis easily, another might have just that capability. But as we've mentioned, once you have learned the logic and rules of one statistical software package, it is usually quite easy to switch to another package.

As explained in Chapter 1, we will be working with the statistical package known as SPSS in this book. In the following, we will describe some of the terminology we will use as we explain the various procedures introduced in the text.

1.7 Terminology used in our instructions

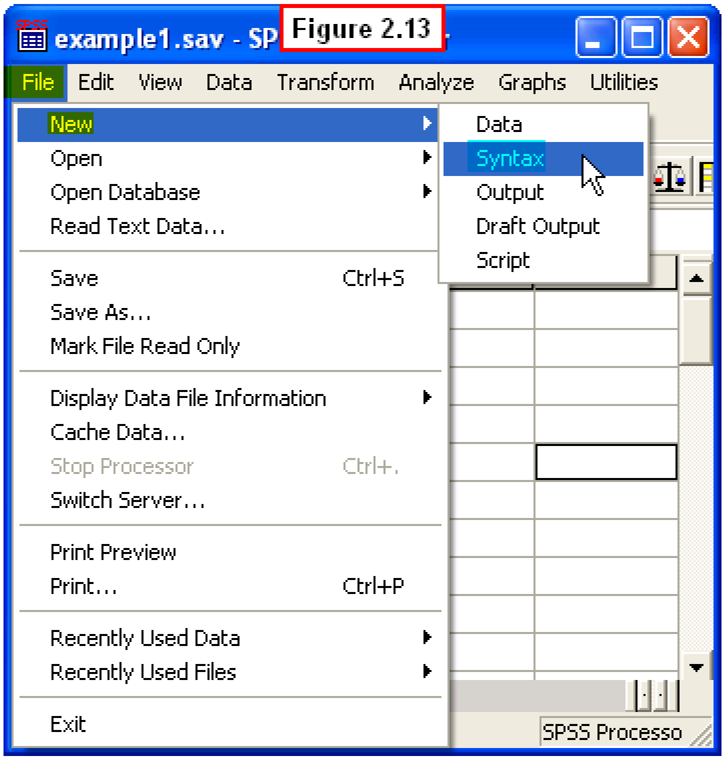
1.7a Understanding our instructions for using SPSS: Drop-Down Menus

To make sure you understand our instructions in each chapter, note the procedure we will use to explain what you need to do as you use what will be called the **Point-and-Click Method** in the examples in subsequent chapters (another method of using SPSS, called the **Syntax Method**, is briefly introduced in Chapter 2).

The point-and-click method simply means that the user moves the onscreen cursor using the computer's mouse to point at various items, then to click an item to perform an operation.
For example, in Chapter 2, we provide the following instructions for opening a new Syntax Editor Window (don't worry about terms like Syntax Editor for now - they are explained in various chapters):

To open a new Syntax Editor Window, select **File, New, Syntax** from the drop-down menu at the top of the Data Editor Window.

We have reproduced Figure 2.13 below to illustrate the above commands:



First, note that **Data Editor Window** is one of three different kinds of windows that the SPSS program uses (more on these in Chapter 2), and is the current window open (or active) on your screen. The above statement is instructing you to open a new SPSS window (called the Syntax Editor Window).

Next, note that the phrase, **File, New, Syntax**, means that you are to use your computer's mouse to click on elements of this Data Editor window. Specifically, we mean that you should move your cursor to point to the word, **File**, at the top of the SPSS Data Editor Window and click the left mouse button. A **drop-down menu** will appear when you click on **File**. This is called a drop-down menu because it drops down from the top of the window when you click on File.

The next thing the phrase, **File, New, Syntax**, indicates is that after you have clicked on the word, **File**, you should then slide your cursor down until the word, **New**, is highlighted and click your left mouse button. When you do that, a new drop-down menu appears, and you should slide your cursor to the right and down until the word, **Syntax**, is highlighted. Last, you should click on the word, **Syntax**. After you have done these three things, the new Syntax Editor will open up in a new window.

Depending on your level of computer experience, the above may sound either very basic or very complicated. If you have worked with computers at all, we are sure that you have followed exactly the same sequence of steps before in, say, opening up a new document window in Microsoft Word. We just wanted to make sure that you understand that when we use a phrase such as select **File, New, Syntax**, that we are essentially saying, "First select and click on 'File,' then select and click on 'New,' then select and click on 'Syntax'.

1.7b Navigating between windows when more than one window is open

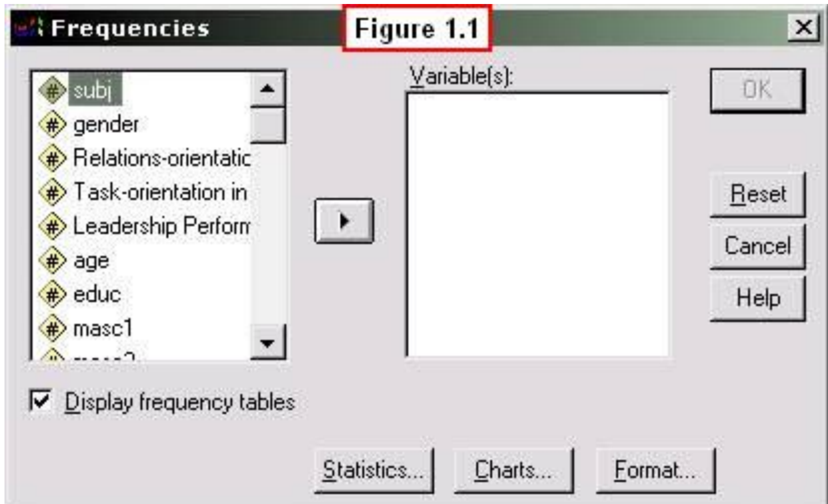
At various points in SPSS procedures, when you click on a selection, a new window will open up in front of the window from which you made your selection (e.g., in the above example, the new Syntax Editor Window opens up, but the original Data Editor Window stays open behind the new window on your computer screen). You will find that there are often several SPSS windows open at the same time.

All currently open windows will be represented with buttons along the **tool bar** at the bottom of your screen (this is the one that has the word, **Start**, on the lower left of your screen). You can navigate back and forth between various open windows by clicking on their respective buttons on the tool bar. Here also, if you have surfed the internet with multiple browser windows open, we are sure you have employed this procedure before.

An alternate way to navigate between windows is to simultaneously press the two keys, **Alt** and **Tab**, on your keyboard. This will cause a little window to open that shows active (i.e., open) windows. You can then select which window you want to be in front of the others by continuing to hold down the **Alt** key while simultaneously tapping the **Tab** key. The window that will be displayed in front will be highlighted, and releasing the **Tab** key will cause that window to move in front of the others.

1.8 Windows, windows, and more windows: Understanding our instructions for using SPSS Dialog Windows

When using the Point-and-Click Method in SPSS, frequently after clicking on a selection, a new window appears on top of the one from which the selection is made. These are called **pop-up dialog windows**, because they pop-up automatically after a selection is made, and because they require some kind of dialog between the user and the software program. This dialog involves the user providing some input that is requested by the program before a procedure can be executed (literally, you have to tell the program what you want it to do by selecting one or more of a list of options the program shows in the pop-up dialog window). An example of such a window is shown in Figure 1.1.



This dialog window requests the user to select which variables s/he wants SPSS to use when it performs the **Frequencies** procedure. In the left panel (sometimes called a pane) of this window is a list of all the variables in the data file that is currently open in SPSS. The right pane is a blank panel labeled **Variable(s):** waiting for input from the user. The user tells SPSS which variables to use in a procedure by moving the desired variables from the left to the right panel.

The little rectangular boxes in this window are actually called **buttons**, because "pushing" them (i.e., clicking on them) causes some action to be performed (just like pushing your power button on your computer causes the action of booting up to be performed). One such button is the **right-arrow** button between the two panels. Clicking it causes the variable that is highlighted in the left panel to be moved over to the empty **Variable(s):** panel on the right. The other buttons in this dialog window (e.g., **OK, Reset, Cancel and Help**) perform the functions indicated by their labels (e.g., clicking on the **Reset** button causes the dialog window to be reset to it's original settings).

Note that the three buttons at the bottom of this window all have three dots (**…**) following their labels. This indicates that if you click on one of these buttons, then a new pop-up dialog window will open requesting additional input. For example, clicking on the **Statistics…** button causes a new pop-up dialog window to appear requesting you to indicate which statistics you want SPSS to compute on the selected variables.

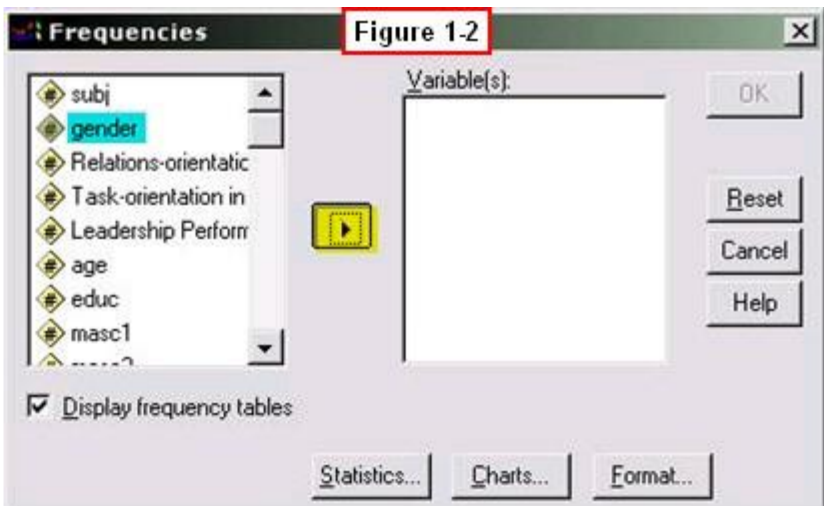
Thus, as we have said, in navigating through a procedure using the Point-and-Click Method, you will often have multiple windows open simultaneously. However, for pop-up dialog windows, once you have made your selections and click the **OK** (or **Continue**) button, the dialogue window automatically closes and returns you to the previous window.

All of the terms we have just described may seem confusing, but once you work with a pop-up dialog window, you will see that the steps are pretty straightforward. In fact, you have undoubtedly worked with pop-up dialog window in other software programs (such as in choosing **Options** in a computer game or responding to input requested on an internet site) – you just didn't know they were called pop-up dialog windows!

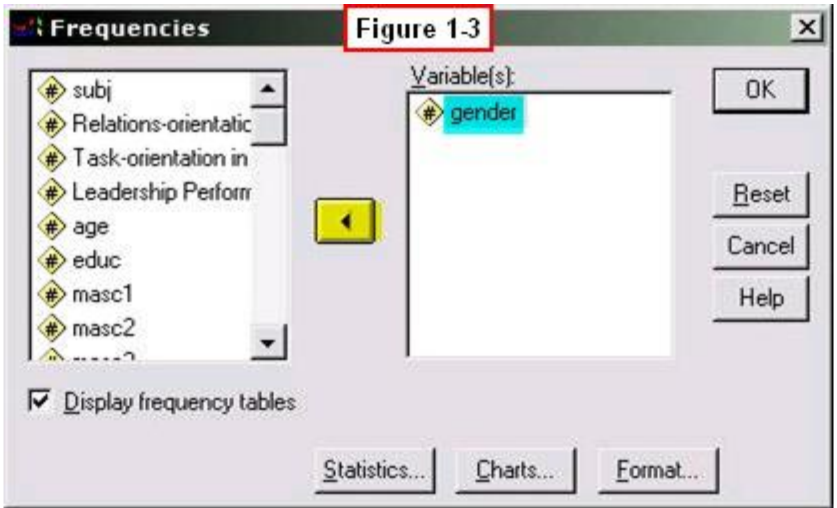
We are simply introducing the terminology and procedures here so that it will be easier for you to understand our instructions in the text chapters. And again, the procedures are fairly intuitive. To return to our example **Frequencies** pop-up dialog window, your input to this dialog request would involve clicking on the variables that you want to select from the left panel and moving them over to the right panel to be analyzed.

As an example, in one chapter we will ask you to select **gender** from the list in the left panel shown in Figure 1.1 in order to have SPSS construct a frequency table of scores on this variable. To do this, we will say: "Highlight **gender** in the left panel and move it to the right panel by clicking the right-arrow button in between the panels."

Figure 1.2 shows what the window would look like after clicking on **gender** to highlight (or select) it, then clicking on the "right-arrow" button in the middle.



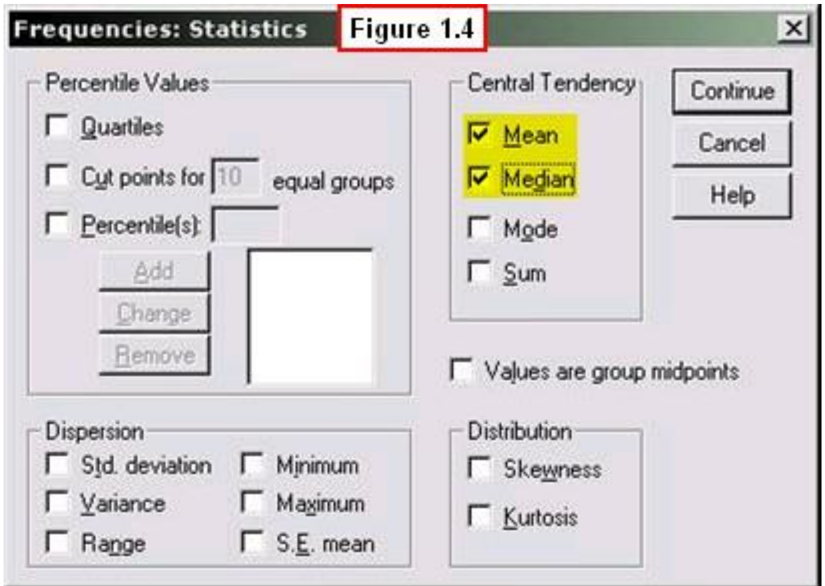
After you have highlighted **gender** and clicked on the right-arrow button, the window will change to look like that shown in Figure 1.3, which now shows gender in the right **Variable(s)** panel.



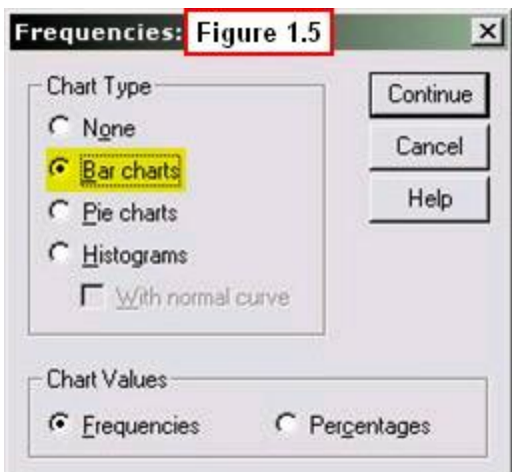
Note that the previous **right-arrow** button in the middle has changed to a **left-arrow** button in Figure 1.3. This is because gender is still highlighted in the right panel, and SPSS allows you to move it back to the left panel if you decide not to include it in the requested procedure. Clicking the left-arrow button would move the highlighted **gender** back to the left panel. If instead, you clicked on **subj** in the left panel, then the button would once again change to a right-arrow to allow you to also move **subj** to the right panel.

A few more terms with examples are provided next. Again, most of these will look familiar to you if you have spent much time on the internet filling out forms – we are simply defining the terms used for these elements now so that you will understand them when we use them in book chapters.

Some pop-up dialog windows ask you to click on **Check boxes** to make selections as inputs. These are blank little white boxes, and a **check mark** appears in the box when clicked. An example of this is shown in Figure 1.4, where we have selected the statistics, **Mean** and **Median**, as the measures of Central Tendency we want SPSS to compute when it performs the Frequencies procedure.



Check boxes are employed by SPSS to allow the user to select more than one item (e.g., both the Mean and Median have been selected in Figure 1.4). Sometimes, however, SPSS requires the user to select only one option from a list of alternatives. In these instances, SPSS employs **Radio Buttons**. Radio buttons are similar to check boxes, except they are blank white circles instead of boxes (see Figure 1.5).

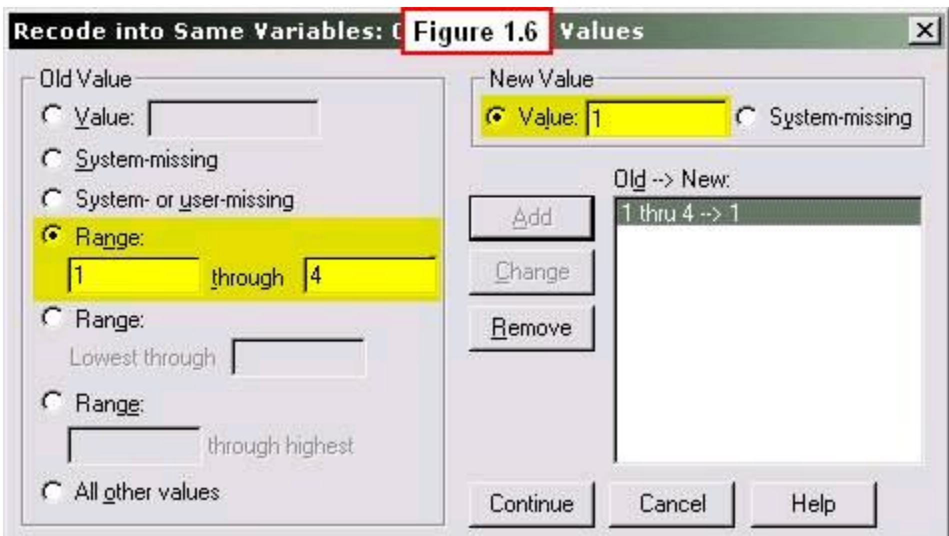


But while the user can click on multiple check boxes, with radio buttons, only one radio button can be selected at any one time in a given panel. For example, in the **Chart Type** panel of Figure 1.5, the **Bar charts** radio button has been selected (a bold dot now appears in the button labeled Bar charts). Once one button has been selected in this panel, if the user clicks on another radio button (e.g., **None**), the new button becomes highlighted, and the previous button returns to a blank white circle.

Thus, you must choose which one selection you want with radio buttons - this should make sense, since SPSS could not create bar charts and no charts at the same time!

Note that the same applies to the **Chart Values** panel at the bottom of the pop-up dialog window - the user must decide whether to select the **Frequencies** radio button or the **Percentages** radio button - both cannot be selected at the same time. The term radio button is used because it is much like pushing buttons on your car's radio - you can only tune in one station at a time!

A final type of pop-up dialogue window used in SPSS involves **text entry boxes** for input. That is, instead of clicking on a radio button or check box to make a selection, the user must actually type in the required information. Here also, you are most likely familiar with these from completing online forms (e.g., where you have to type in your e-mail address into a blank box). In most cases, the "text" that we will enter into these boxes will consist of digits, as shown in Figure 1.6.



In the **Old Value** panel on the left of the **Recode into Same Variables:** pop-up dialog window shown in Figure 1.6, we clicked on the **Range:** radio button that originally had blank **text entry boxes** below it. We then typed in the digits **1** and **4** to indicate this is the range of old values to be recoded into the **New Value** panel in the upper right. We typed a **1** in this text entry box to indicate that all scores from 1-4 (the old values) should be recoded to 1 (the new value).

Don't worry if you don't understand the Recode procedure described above - it will be explained in much greater detail in Chapter 5. All that is important now is that you understand that when we use the term, text entry box, this indicates that you will have to type something into the box.

1.9 Final Reassurances & The SPSS Help Features

If you've worked with computers before, much of what has been presented here has been review. But if your use of computers has been limited, you may feel overwhelmed by all the information. Three points you should keep in mind:

- It takes time to learn about the many different aspects of computing.
- As you gain practical experience at the computer, you will have a better framework for the ideas and information that we have been discussing.
- The startup time invested in this learning will yield great dividends in time and effort in the future.

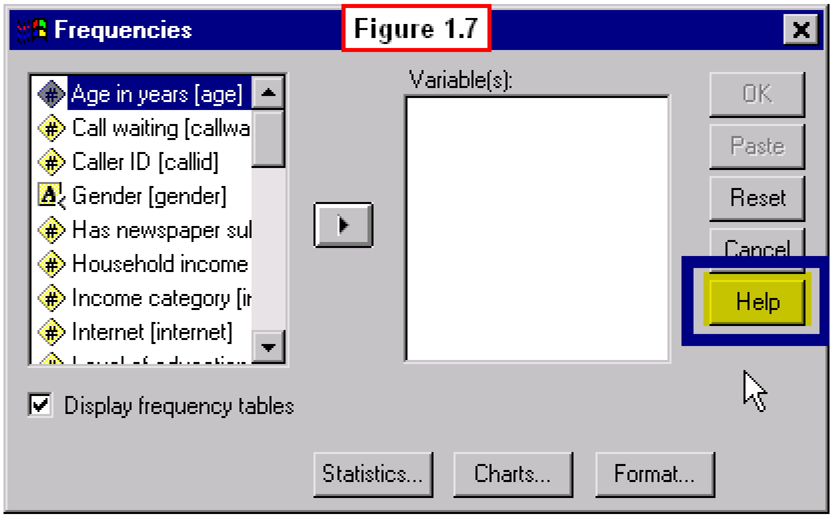
In the meantime, you may want to review the terms in this chapter that are bold-faced, or reread sections after your first session at the computer terminal. You may be surprised to learn how easy it is to use the computer for the purpose of analyzing data. It's merely a matter of acquiring some basic rules, of becoming accustomed to computer jargon, and of overcoming the anxiety that comes with having to learn a new software application that you initially don't understand. Be assured, this anxiety will diminish as you gain hands-on experience.

Perhaps most important of all, don’t worry about losing files, making errors, or doing other things that might alter or "ruin" your file. This book is set up so that no catastrophes can occur - every ruined or lost file can be re-established in a matter of minutes, every problem can be easily fixed, and every procedure can be redone if it is not completed correctly the first time. Of course, you will save time and effort if you follow the directions in the chapters carefully.

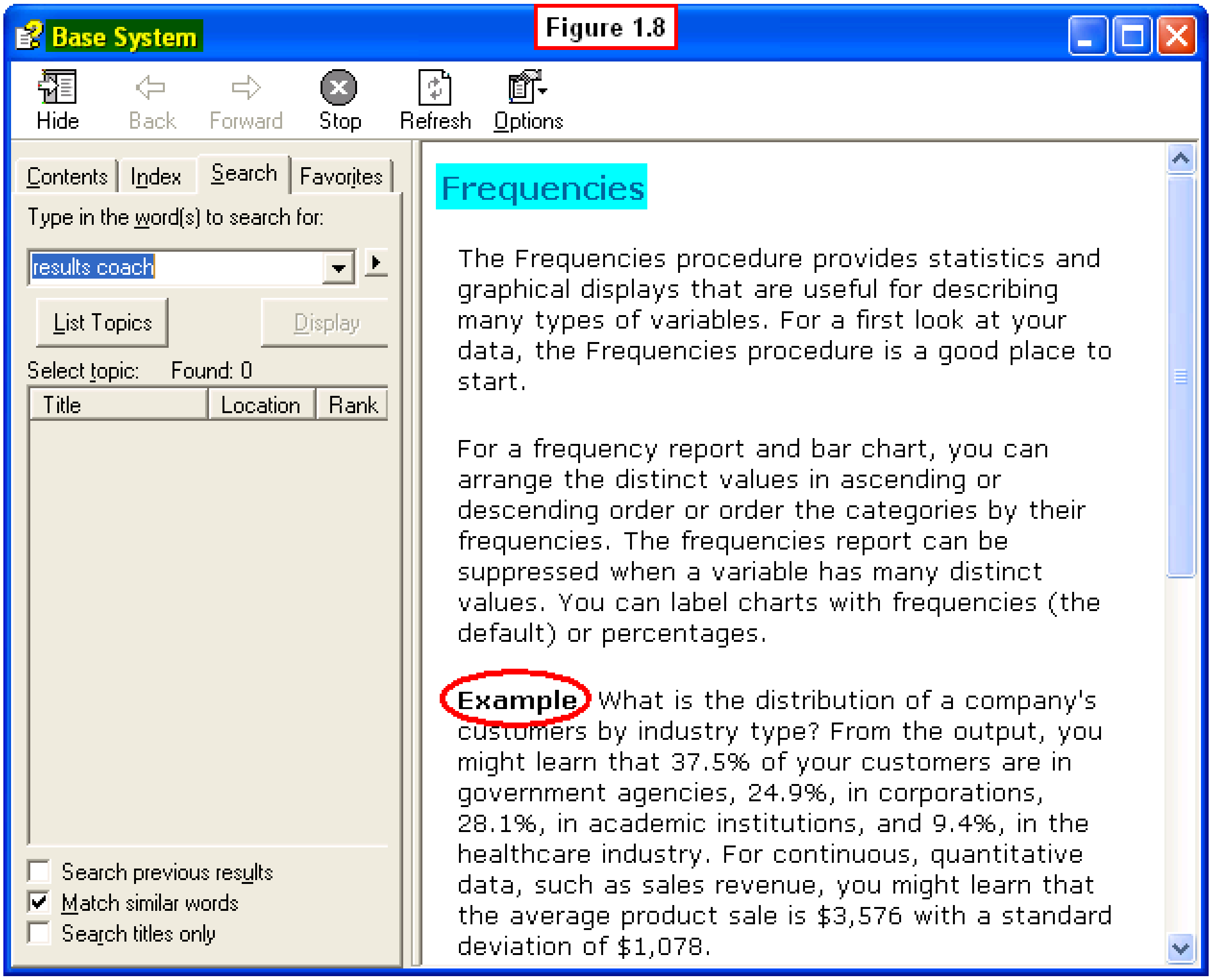
But even though it may take substantial time to learn the procedures the first time around, if you make major errors, you can typically retrace your steps and re-establish your starting point within a matter of minutes. In the remainder of this appendix, we introduce various ways and places you can access the many SPSS help features from within the program.

1.9a Dialog window help

Although we have endeavored to fully explain all important features of dialog windows for the procedures introduced in this text, there may be times when you forget or become confused when a new dialog window appears. Or you may be just curious about options that we do not discuss. You can obtain more information simply by clicking the Help button (Figure 1.7).



Clicking the Help button opens up a new Base System Help window providing general information about the procedure in the dialog window, along with an example and other helpful information. Figure 1.8 shows the window that opens when the Help button in Figure 1.7 is clicked.



1.9b Context Menu help

You can quickly obtain help about any item in a dialog box (and generally anywhere else in SPSS) by moving your mouse on top of the item and pressing the right button on your mouse (also referred to as **right-clicking** on the item). This causes a **Context Menu** to be activated, so-called because they provide help from anywhere in the context of the window.

In Figure 1.9 the user has right-clicked on the word, *Mode,* to obtain help. A white text box appears with an explanation of the item that has been selected. Again, context menu help is available in most places in SPSS.

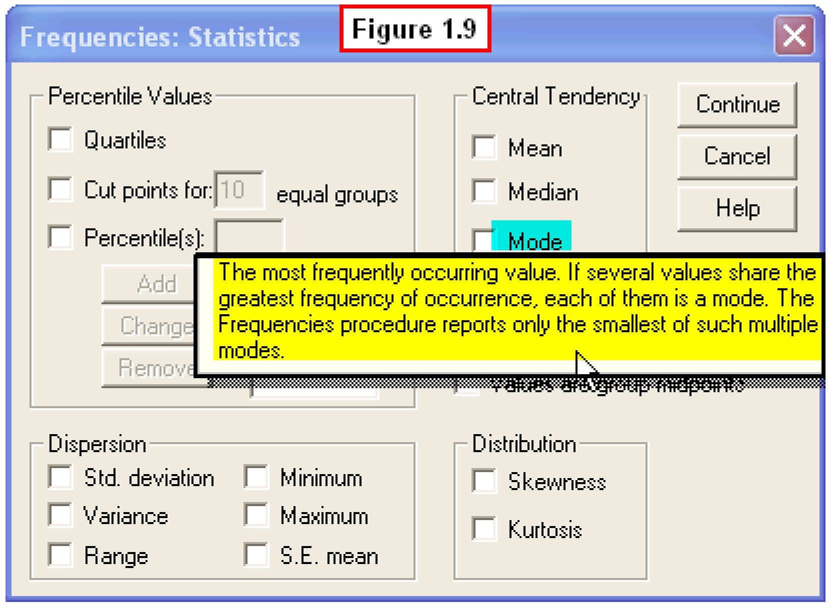
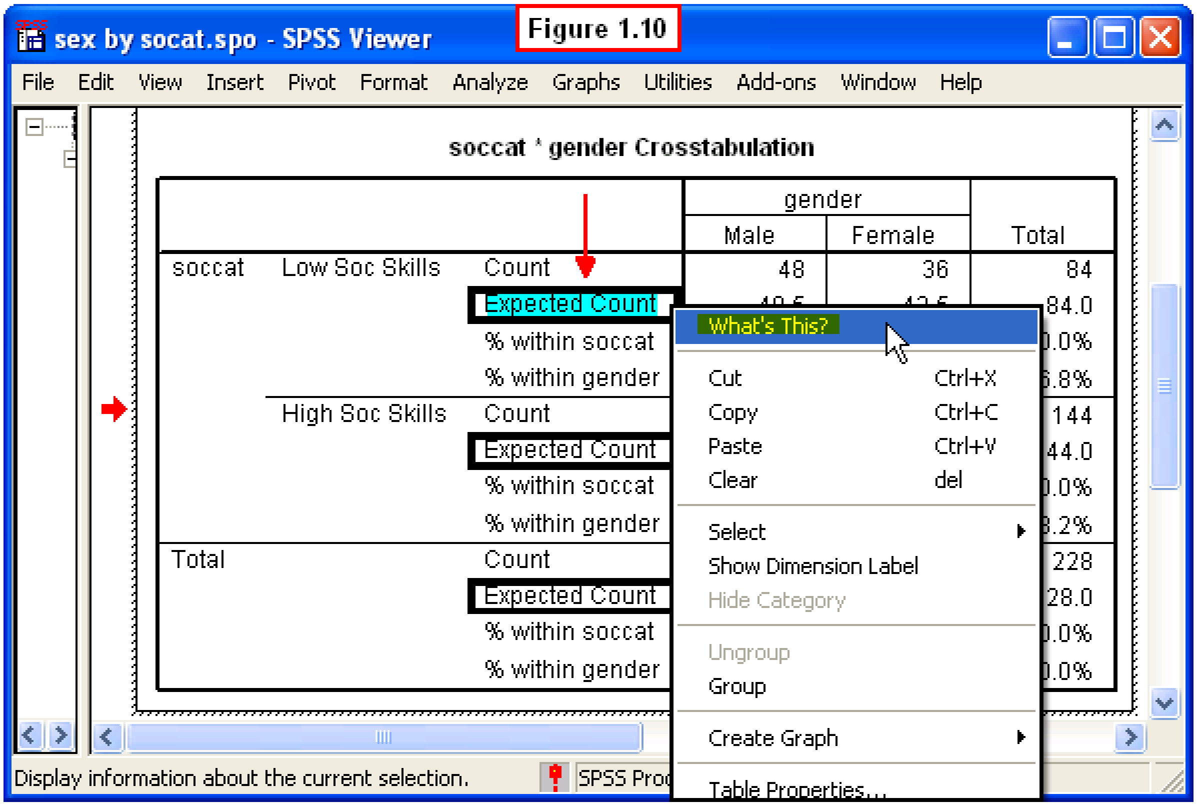


Figure 1.9

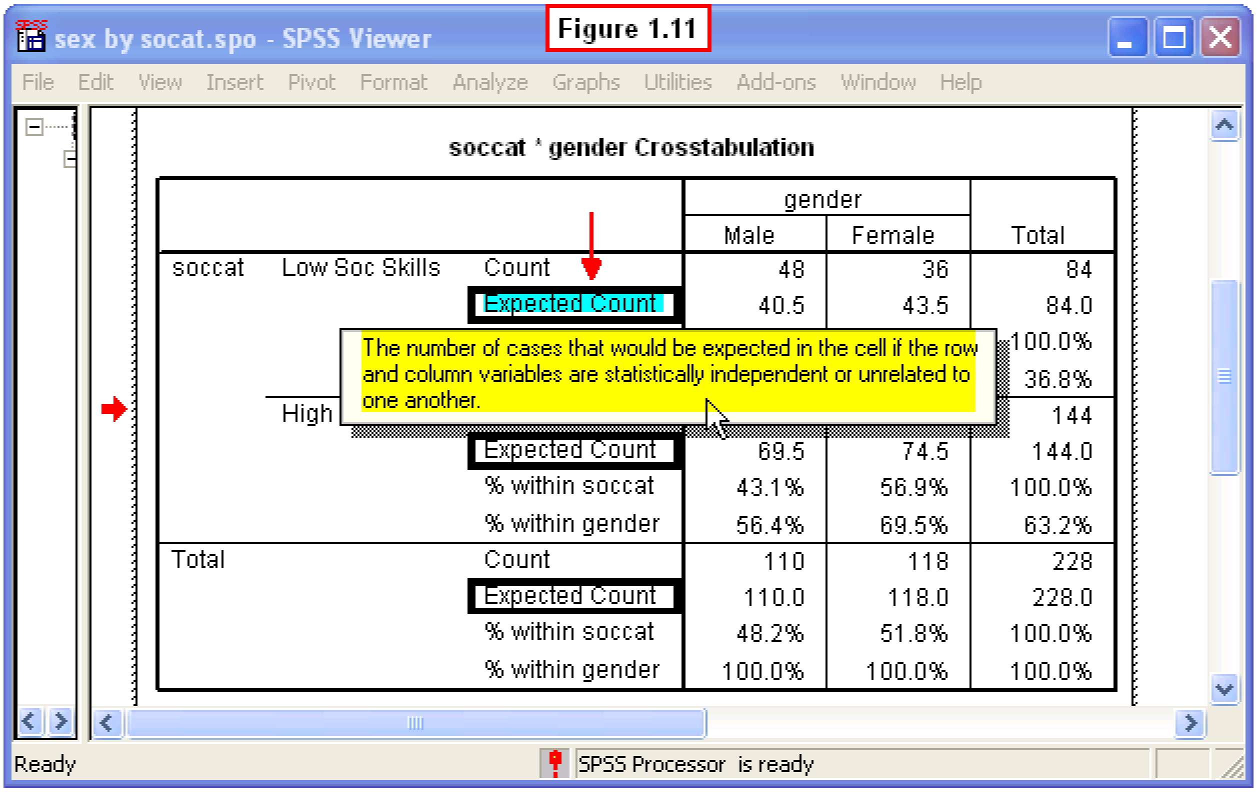
- **Note about Left and Right Mouse Button Clicks:** in the video below, **red circles** around the cursor indicate a **left-click**, while **blue circles** indicate a **right-click** of the mouse.

1.9c Help within output tables

SPSS uses yet another type of window to display results of statistical analyses (called an **SPSS Viewer** window). If you are confused about any item in the output file, you can use the context menu help in this window also. Figure 1.10 gives an example of such context help menu in an output file of the results of a crosstabulation procedure. The user has right-clicked the term, *Expected Count*, then left-clicked *What's this?*



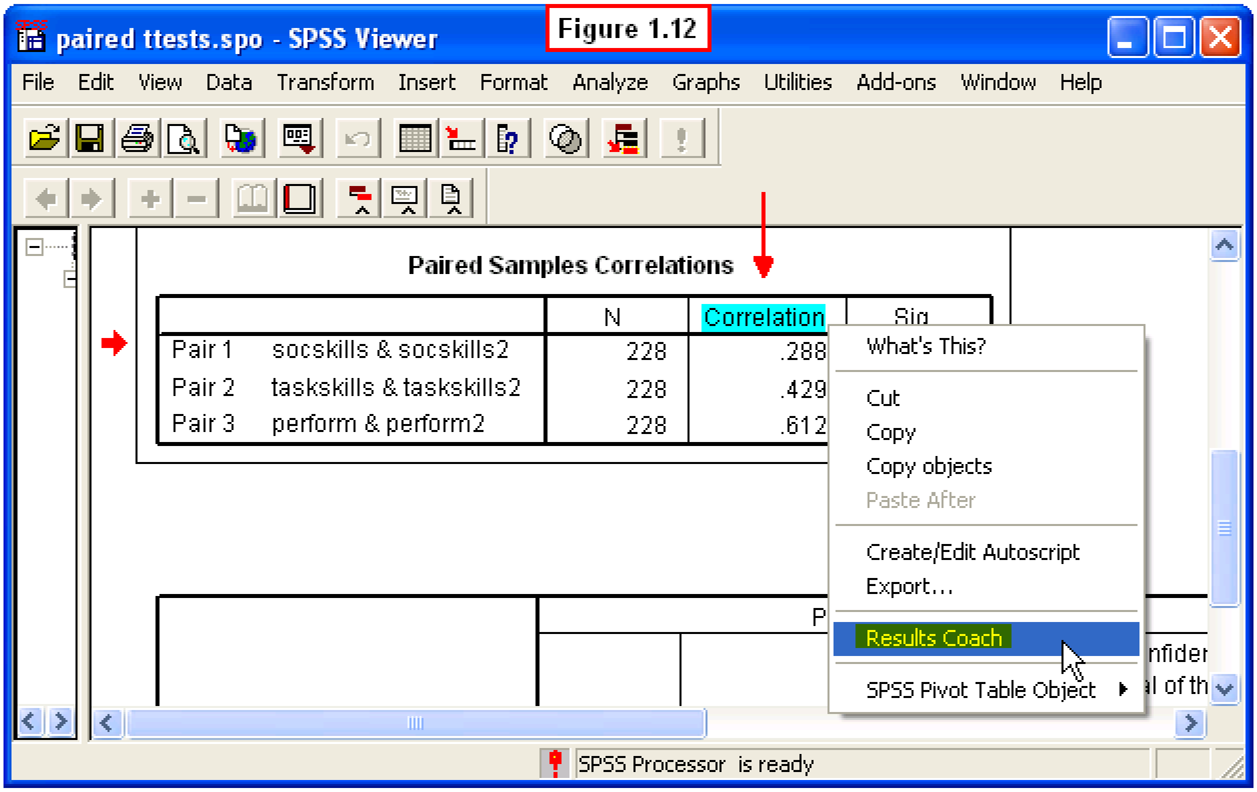
A text box appears giving an explanation of the term *Expected Count* (Figure 1.11). These boxes can be very helpful if you are having trouble understanding any part of the output of an analysis.



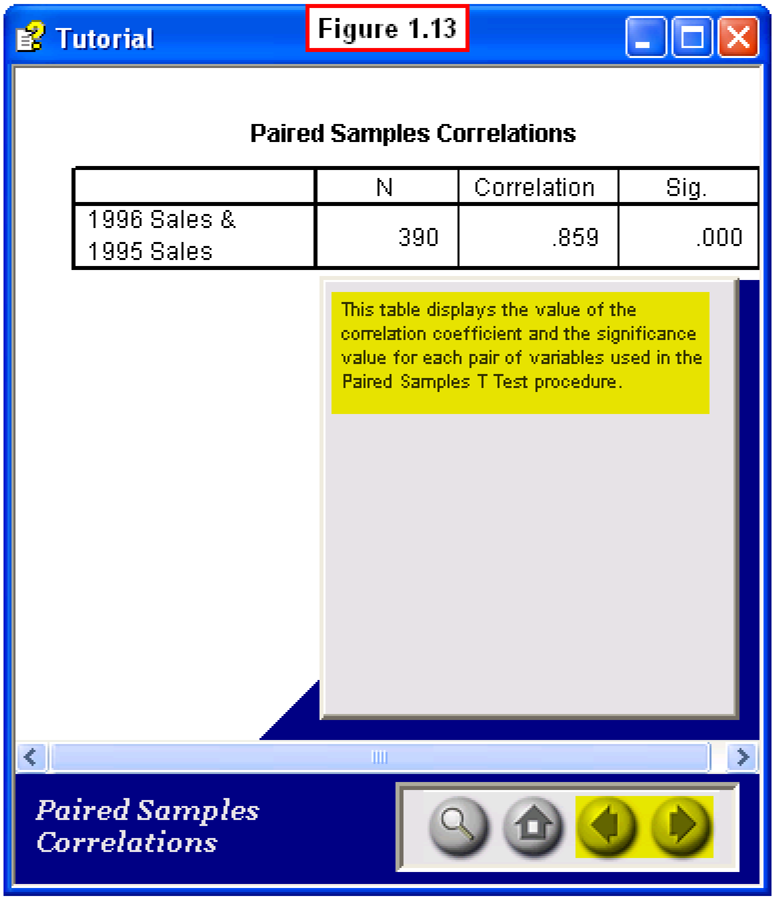
- **Note about Left and Right Mouse Button Clicks:** in the video below, **red circles** around the cursor indicate a **left-click**, while **blue circles** indicate a **right-click** of the mouse.

1.9d The Results Coach

SPSS provides coaches that take you step-by-step through the process of interpreting results or deciding which statistical analyses to choose by providing helpful examples. In Figure 1.12, the user has right-clicked the word, *Correlation*, and left-clicked the **Results Coach** option to get help with this term.



An SPSS Tutorial window will then appear (Figure 1.13). Clicking the right-arrow button in this window will take the reader through a series of explanations of this statistic. As we have mentioned, in this text we provide detailed explanations of outputs from the examples used, but the Results Coach is always there if you forget or get confused.

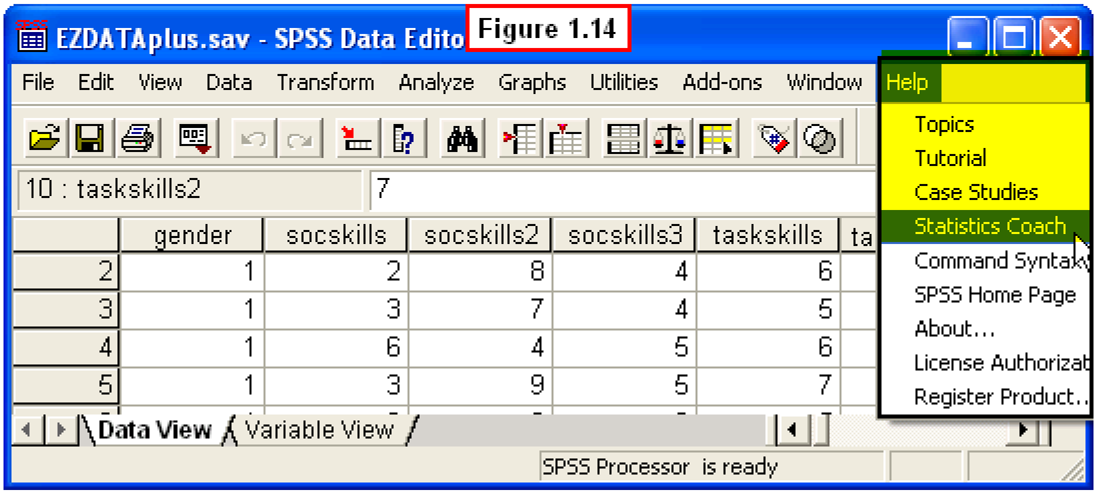


- **Note about Left and Right Mouse Button Clicks:** in the video below, **red circles** around the cursor indicate a **left-click**, while **blue circles** indicate a **right-click** of the mouse.

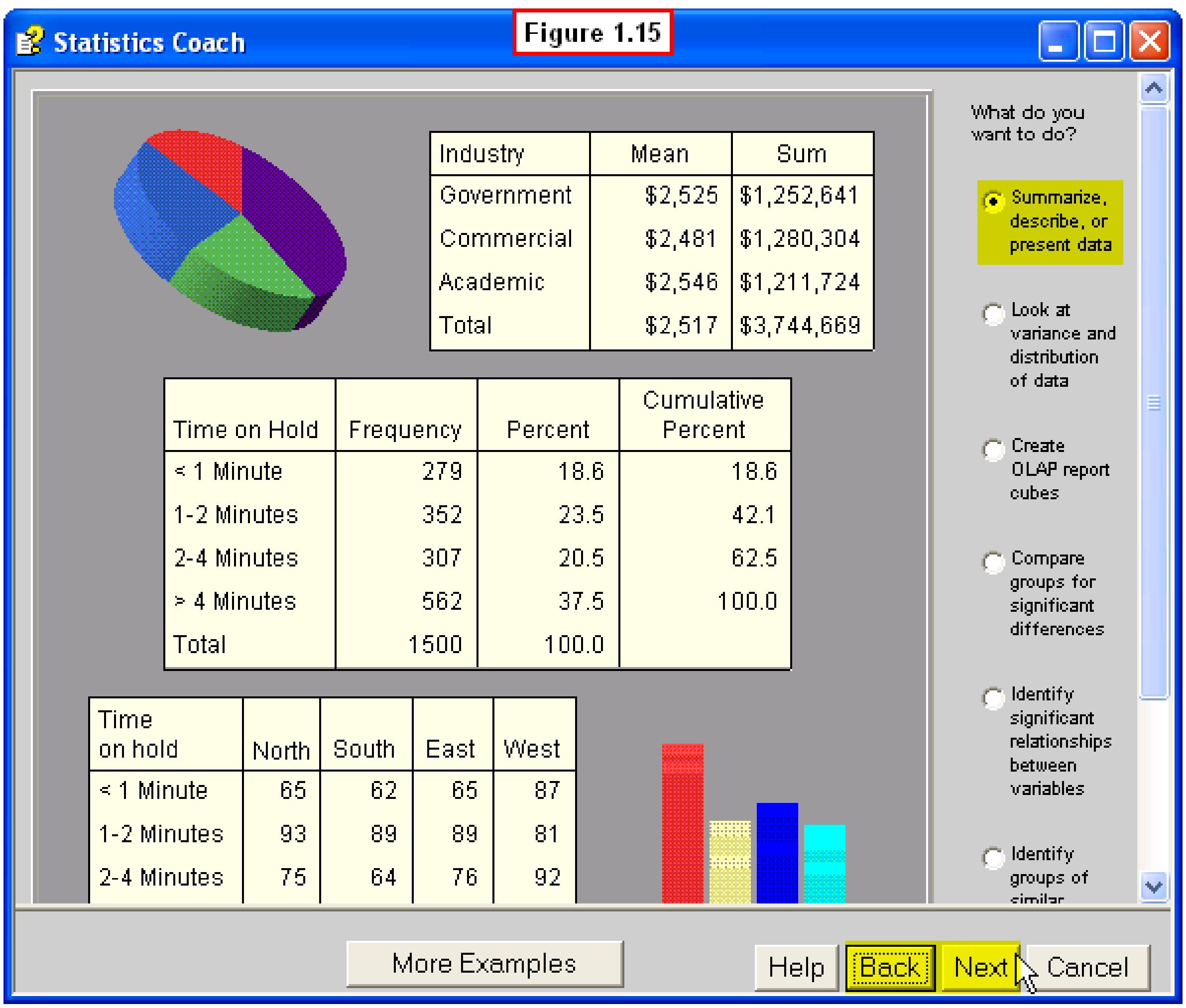
1.9 e The Statistics Coach

In all the examples in this book, we specify what analyses to run. But if you are curious or have questions about what statistical procedure is appropriate to use on some future data set, you can also use the **Statistics Coach**.

Figure 1.14 shows how to access the Statistics Coach. The user has left-clicked the **Help** drop-down menu in the Data Editor window, then clicked on the **Statistics Coach** menu option.



An SPSS Help dialog window then appears (Figure 1.15) In this figure, the user has selected *Summarize, describe...* in response to the question "What do you want to do?



1.10 Exercise

Take some time to make certain you understand the terms listed below. If you need to, refer back to the section of the appendix where the concept is discussed.

Bivariate - 1.3
Buttons - 1.8
Check Boxes - 1.8
Context Menu Help - 1.9b
Correlation - 1.3
Data - 1.5
Descriptive - 1.3
Dialog Windows - 1.8
Dialog Window Help - 1.9a
Drop-Down Menus - 1.7
Experimental Method - 1.4
Information - 1.5
Multivariate - 1.3
Output Table Help - 1.9c
Radio Buttons - 1.8
Results Coach - 1.9d
Software - 1.6
Statistical Package - 1.6
Statistics Coach - 1.9e
Text Entry Boxes - 1.8
Univariate - 1.3

Review Videos

1.1
http://youtu.be/7jNyU5AYOjE

1.2
http://youtu.be/-jkY8HTXhvU

2.1
http://youtu.be/n29cbZSZgQc

2.2
https://www.youtube.com/watch?v=Q_J9quQSnCg

2.3
https://www.youtube.com/watch?v=kw5n7pTH-40

2.4
https://www.youtube.com/watch?v=M88TETwdT0o

2.5
https://www.youtube.com/watch?v=Px3gsHTcsLc

2.6
https://www.youtube.com/watch?v=TnWx60TOShI

2.7
https://www.youtube.com/watch?v=qhZV0vQe_3Y

2.8
https://www.youtube.com/watch?v=UGV-by04Q0Y

2.9
https://www.youtube.com/watch?v=-Uyy_a7NDPg

3.1
http://youtu.be/bW1d5b2jijk

3.2
http://youtu.be/aC0Dkz9HOIo

3.3
http://youtu.be/KeoANURRJ4I

5.1
http://youtu.be/dV50uUrtV_Y

5.2
http://youtu.be/3EbYDse7hHY

5.3
https://www.youtube.com/watch?v=xfPWZYyuusY

5.4
http://youtu.be/uXcpZ_oUCOM

7.1
https://www.youtube.com/watch?v=WVCvsCyNenc

8.1
https://www.youtube.com/watch?v=d35TXLw0o24

9.1
http://youtu.be/6Z82YthlT7o

10.1
http://youtu.be/p5eh-HxtMIo

11.1
http://youtu.be/HfaX_Nvek64

12.1
http://youtu.be/v9ab8KAa-2k

13.1
http://youtu.be/sxWSAl30P9I

14.1
http://youtu.be/JnfCGgUs9eA