

SPSS 15.0 Supplement for: Biostatistics for the
Health Sciences

R. Clifford Blair

and

Richard A. Taylor

February 5, 2008

© R. Clifford Blair and Richard A. Taylor 2005

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of R. Clifford Blair or Richard A. Taylor.

Preface

This manual is not an introduction to SPSS[®] and contains only a portion of the material that would be contained therein.¹ Rather, this manual teaches the reader how to use SPSS 15.0 for Windows to analyze data related to the problems addressed in *Biostatistics For The Health Sciences* by R. Clifford Blair and Richard A. Taylor.

To this end, SPSS is used to analyze various of the data sets and problems contained in that text. The reader will be required, especially in the early part of the manual, to input selected data to the system. Additional data sets accompany this manual so that the reader will not be burdened with data entry once this skill is acquired.

It is assumed that the reader is familiar with the Microsoft Windows[®] operating system so that little instruction in this regard will be provided. It is further assumed that the reader is using SPSS 15.0 rather than some earlier version.

Chapter 1 familiarizes the reader with some basic SPSS concepts. Key among these is the entry and saving of data which is prerequisite to all that follows. Chapters 2 through 10 address the corresponding chapters in the text. That is, Chapter 2 in the manual addresses problems in Chapter 2 in the text and so on. Section headings in these latter chapters reflect the sections/pages in the text addressed in that section of the manual. Thus, while studying the text readers can quickly locate the comparable area in the manual by referring to the manual's table of contents. In addition, a manual appendix relates specific examples, data tables etc. in the text to specific pages in the manual. These provisions allow easy integration of SPSS analyses as study of the text progresses.

Finally, we would caution that data analysis, whether by hand calculations or software is no substitute for a clear understanding of the underlying statistical concept. Such analysis can, however, be an invaluable aid to mastery of such concepts and is indispensable for research purposes.

¹Such introductions to the SPSS[®] analysis system can be found in references [1] and [2].

Contents

1 Preliminaries	1
1.1 Introduction	1
1.2 Getting Started	1
1.3 Inputting Data	2
1.4 Performing Analyses	6
1.5 Editing and Printing Your Output	8
1.6 Creating New Variables	9
1.7 Graphs	11
Exercises	12
2 Descriptive Methods	13
2.1 Introduction	13
2.4 Distributions	13
2.5 Graphs	19
2.6 Numerical Methods	28
Exercises	34
3 Probability	35
3.1 Introduction	35
3.3 Contingency Tables	35
3.4 The Normal Curve	36
Exercises	38
4 Introduction to Inference and One Sample Methods	39
4.1 Introduction	39
4.2 Sampling Distributions	39
4.3 Hypothesis Testing	41
4.4 Confidence Intervals	47
Exercises	48
5 Paired Samples Methods	49
5.1 Introduction	49
5.2 Methods Related to Mean Difference	49
5.3 Methods Related to Proportions	51

5.4	Methods Related to Paired Samples Risk Ratios	53
5.5	Methods Related to Paired Samples Odds Ratios	54
	Exercises	56
6	Two Independent Samples Methods	57
6.1	Introduction	57
6.2	Methods Related to Differences Between Means	57
6.3	Methods Related to Proportions	60
6.4	Methods Related to Independent Samples Risk Ratios	61
6.5	Independent Samples Odds Ratios	61
	Exercises	62
7	Multi-Sample Methods	63
7.1	Introduction	63
7.2	The One-Way Analysis of Variance (ANOVA) F Test	63
7.3	The 2 By k Chi-Square Test	64
7.4	Multiple Comparison Procedures	65
	Exercises	66
8	The Assessment of Relationships	67
8.1	Introduction	67
8.2	The Pearson Product-Moment Correlation Coefficient	67
8.3	The Chi-Square Test for Independence	68
	Exercises	69
9	Linear Regression	71
9.1	Introduction	71
9.2	Simple Linear Regression	71
9.3	Multiple Linear Regression	72
	Exercises	74
10	Methods Based on the Permutation Principle	75
10.1	Introduction	75
10.3	Applications	75
	Exercises	78

Chapter 1

Preliminaries


1.1 Introduction

The hand calculations you perform in conjunction with your studies of the statistical concepts in *Biostatistics For The Health Sciences* by R. Clifford Blair and Richard A. Taylor¹ are designed to enhance your understanding of the concepts under study. Modern statistical practice only rarely resorts to such methods for the analysis of data collected for research purposes. This is because (1) data sets collected in applied research contexts are often too large for hand calculations and (2) such calculations are prone to errors. Instead, researchers resort to tried and true statistical software that can be relied upon to provide quick, accurate results. The SPSS system is one of the most mature and popular of these.

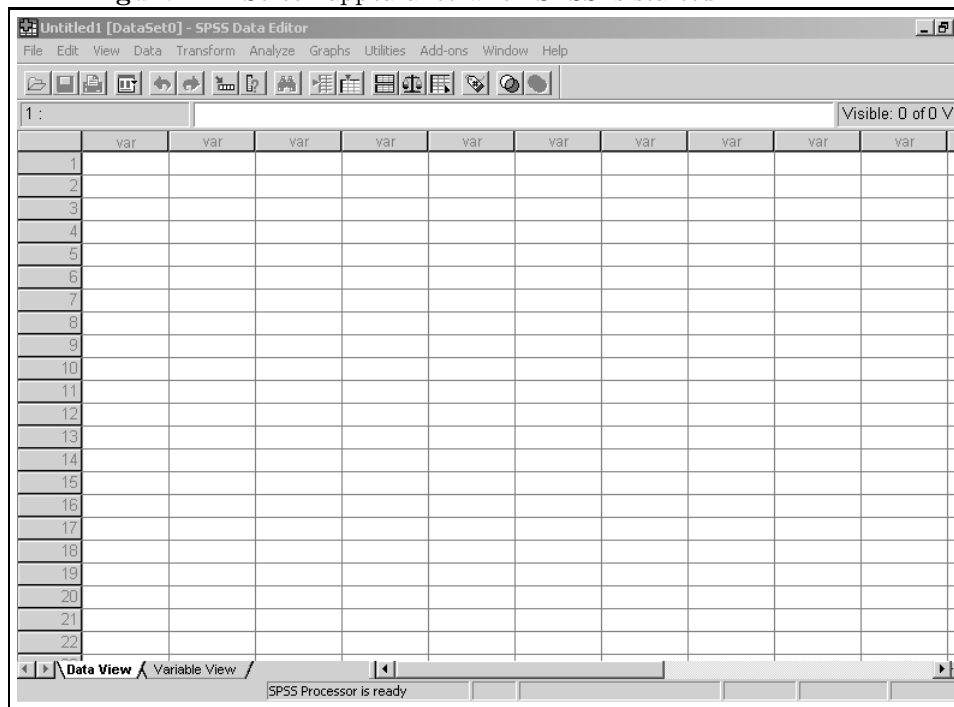
In this manual, you will learn to use SPSS to analyze data associated with the problems outlined in the text. You will learn some basic SPSS concepts in this chapter. Chapters 2 through 10 will address problems in the like numbered chapters in the text.

1.2 Getting Started

Before starting SPSS, you should make yourself a directory in which to store your work. I've called my directory `bhs_dir`. You should then copy the files that accompany this manual into your directory.

To start SPSS you can either double click the SPSS shortcut icon on the desktop  if you have one, or select the sequence **Start⇒Programs⇒SPSS for Windows⇒SPSS 15.0 for Windows**. The notation **Start⇒Programs⇒SPSS for Windows⇒SPSS 15.0 for Windows** means that you are to first select the Start menu, then Programs, then SPSS for Windows and finally SPSS 15.0 for Windows. We will use this notation throughout this manual. Unless some

¹Henceforth referred to as “the text”.

Figure 1.1: Screen appearance when SPSS is started.

of the default settings for SPSS have been altered, the windows shown in Figure 1.1 will appear. If the insert panel shown in Figure 1.2 appears, put a check in the **Don't show this dialog in the future** box and click **OK**. We won't need this panel.

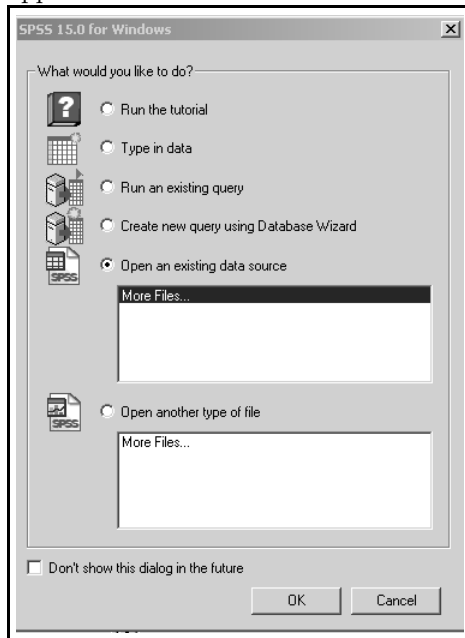
Returning to the opening screen depicted in Figure 1.1, notice the two tabs at the bottom of the screen. The **Data View** tab brings up the Data Editor that is used for the input of data. This is probably the screen that was active when you started SPSS. But to use data properly, SPSS must have a description of that data. Clicking the **Variable View** tab brings up the window that allows for this description. You will gain experience in the use of these two screens as we proceed through this manual.

1.3 Inputting Data

Logically, before you can analyze data, you must input it to the analysis system. This is done by either typing the data into the Data Editor or loading a preexisting data set. You will learn how to load a stored data set shortly. For now, let's see how data is entered from the keyboard. We'll use the data in Table 2.6 on page 34 of the text for this demonstration.

Entering data from the keyboard is a two-step process. First we describe

Figure 1.2: Unneeded screen that may appear when SPSS is started.

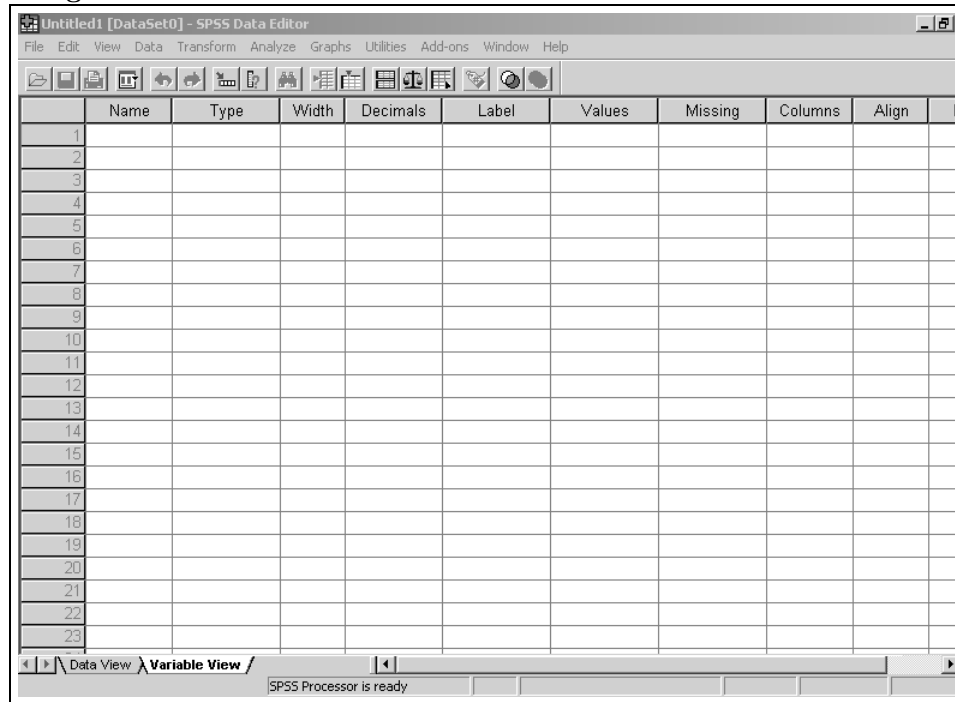
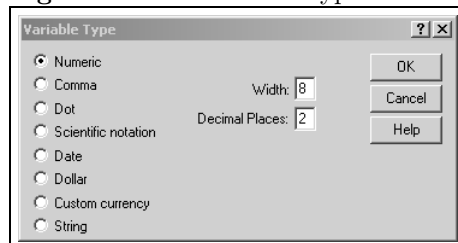


the data to SPSS then we enter the variable values. To describe the data click the **Variable View** tab. This brings up the screen shown in Figure 1.3.

Each row in this window is used to describe a single variable. We have only one variable in Table 2.6 (i.e. x) so we will need only the first row for this example. The first column is headed **Name**. This is where we name the variable. We will name our variable x because that was the only designation used in table 2.6 but we would usually want to use a more informative name such as “gender”, “blood_pressure” or “age.” Variable names must begin with an upper or lower case letter and contain no blanks. That is why we used the underscore with the name ‘blood_pressure.’ Type an x in the cell just below the word **Name**.² Press **Enter** or just click in the cell under the word **Type**.

In either case, SPSS, with the exception of the **Label** column, fills the cells in the remainder of the row with default values. Notice that the default for **Type** is *Numeric*. This is the option we will use most often. It indicates that standard numbers in almost any recognizable format are to be entered. To see other options, click the button with three dots in the corner of the cell. The **Variable Type** panel shown in Figure 1.4 appears. The options for **Type** are shown on the left side of this panel. For example, if we wished to type values with commas setting off groups of three digits such as 1,333,162 we would choose the **Comma** option. As you can also see, there are options for entering dates,

²We will use italics to indicate things you are to type or choose in windows.

Figure 1.3: Variable view.**Figure 1.4:** The variable type window.

dollars etc. We will employ a few of these in later chapters.

Because the numbers in Table 2.6 are single digits with no decimal places, we will set the number of Decimals to zero. We can do this by typing a 0 (zero) in the **Decimal Places:** box of the **Variable Type** panel. Now type a 1 in the **Width:** box to show that the data has total width of one and click **OK** to save your changes and close the panel. Setting the **Decimal Places:** and **Width:** options permits SPSS to more attractively format output from various procedures as well as making the output more intelligible.

Notice that the **Width** and **Decimal** columns in the **Variable View** reflect the changes we made to the **Decimal Places:** and **Width:** options in the **Variable Type** panel. Had we chosen to do so, we could have entered our changes in **Variable View** without resort to the **Variable Type** panel by simply clicking in the cell where a change is to be made and using the up and down scroll device. Changes in the **Variable Type** panel will be reflected in **Variable View** and visa versa.

We will not comment further on the columns in **Variable View** at this point except to note that the column headed **Measure** deals with the scale of measurement as discussed in Chapter 2 of the text. You have three choices for this option, scale, which indicates that the values are associated with an interval or ratio scale, or you may choose nominal or ordinal which reflect the scales with those names discussed in the text. Leave this option as *scale* which is the default.

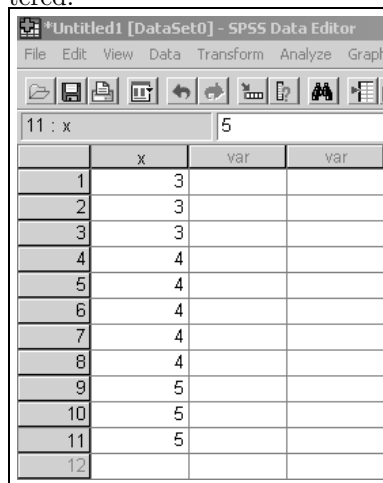
With the characteristics of our data defined, we now begin the data entry process. To do this, click the **Variable View** tab to open the Data Editor. Notice that the first column is headed **x** which is the name we gave our variable. All other columns are headed **var** which is the default. Had we defined a second variable in **Variable View**, its name would appear in the second column, a third variable name would appear in the third column and so on.

The form of the Data Editor will be familiar to most readers as it is similar to a spreadsheet. Data entry is straightforward. To begin entering the data from Table 2.6 click in the first cell of the first column (i.e. the first cell below the **x** heading). Type the number 3 which is the first number in Table 2.6. Move down to the next cell and type another 3 which is the second number in the table. Continue until all data are entered. Your Data Editor screen should appear as shown in Figure 1.5.


It is wise to save ones work at frequent intervals so let's save our data at this point. To do so click **File⇒Save As....** When the **Save Data As** window opens navigate to the directory you created earlier (if necessary) and enter *table 2.6 data* in the **File name:** window. Be sure that *SPSS (*.sav)* appears in the **Save as type:** window and click *Save*. The .sav file type places the file in the SPSS native format. As you can see, there are many other choices for file formats. This data set is now available to us at any time we choose to load it into the Data Editor.


To see that this is true, let's clear the Data Editor and reload the data from the disk directory. To clear the current data from the editor click **File⇒New⇒Data**. This opens a new data sheet and places it over the old one. Now activate (i.e.

Figure 1.5: Screen appearance when data from Table 2.6 are entered.



	x	var	var
1	3		
2	3		
3	3		
4	4		
5	4		
6	4		
7	4		
8	4		
9	5		
10	5		
11	5		
12			

click on) the old (original) data sheet. (You may have to use the **Restore Down** button  to reduce the window size so you can access the original data sheet.) This directs the attention of SPSS to the old sheet. Now click **File⇒Close**. If SPSS asks you if the contents of the Data Editor should be saved click *No*. This closes the data sheet.

Now let's reload it from the saved copy. Click **File⇒Open⇒Data....** (Or click the open file icon .) Click on **table 2.6 data.sav** then click **Open**. Your data is restored to the **Data Editor**.

1.4 Performing Analyses

Now that we have entered our data, we can use SPSS to perform analyses on these data. We will provide a simple example in this section but will give numerous examples in the chapters that follow.

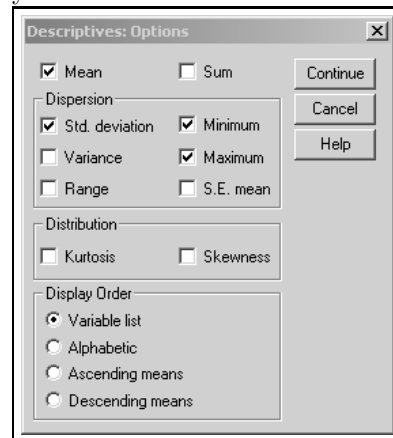
SPSS uses two basic methods for performing analyses and other tasks. The first is menu driven while the second makes use of one of several programming languages. We will cover only the use of menus for analyses in this manual. To learn other methods, consult reference [1].


Before beginning the analysis, be sure your data are displayed in the Data Editor. If they are not displayed,³ select **File⇒Open⇒Data...**, navigate to the directory in which you saved **table 2.6 data** and open it.

Let's begin by calculating a few descriptive statistics on the variable we've called *x*. To do this, select **Analyze⇒Descriptive Statistics⇒Descriptives....**

³Your data will not be displayed if you've ended the SPSS session or closed the worksheet.

Figure 1.6: Screen appearance for options selected for the analysis of the Table 2.6 data.

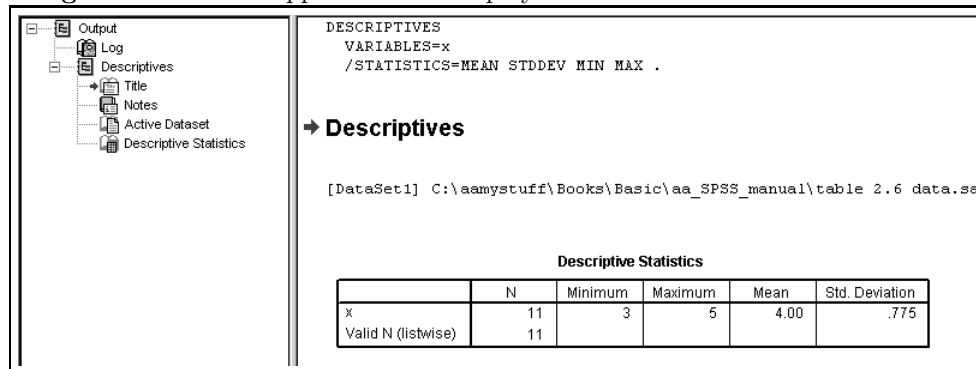


The **Descriptives** window opens. Notice that the variable *x* is highlighted in the left hand window. If it isn't highlighted, select it by clicking on it. Now move it to the **Variables(s):** window by clicking the  button. In general, all variables in the data set will be listed in the left hand window and you will move those on which you wish to perform analyses to the right hand window as you did with *x*.

You can determine which statistics will be displayed by clicking the **Options...** button. If you use this option, a window opens listing all the statistics that may be displayed with this command. Those that are checked will be displayed. Be sure that checks are placed by **Mean**, **Std. deviation**, **Minimum** and **Maximum**. Darken the **Variable list** option. Your screen should appear as shown in Figure 1.6. Now click **Continue** and then **OK**. A **Viewer Window** opens displaying the content shown in Figure 1.7.

As may be seen, the **Viewer Window** is divided into two sections, or as SPSS refers to them, panes. The lefthand pane is called the **Outline Pane** while the righthand pane is termed the **Display Pane**. You can resize these panes by placing the pointer on the vertical gray bar that separates the two. When the symbol appears click and drag the bar to the left or right. We will make limited use of the **Outline Pane** later but for the moment will concentrate on the **Display Pane**.

Earlier we indicated that commands can be issued to SPSS by clicking various menu items or by use of one of several scripting languages. We are using the menu approach in this manual but SPSS implements our menu driven commands by writing the appropriate scripting language commands. This is what you see as the first three lines at the top of the **Output Page**. It is good practice to briefly review these lines when you produce output as this practice

Figure 1.7: Screen appearance for display of some basic statistics.

aids in learning the scripting language without too much effort.

The line

```
[DataSet1] C:\aamystuff\Books\Basic\aa_SPSS_manual\table 2.6 data.sav
```

identifies the source of the data used in the analysis. This line in your output will be of similar form but will not be identical to ours.

The results of our analysis appears in the table under the heading **Descriptive Statistics**. As may be seen in this table, the variable analyzed is named *x*, there are 11 valid observations in the data set the minimum and maximum values of which are three and five respectively. The mean and standard deviation of the 11 observations are 4.0 and .775 respectively. You will learn about the mean and standard deviation as well as other descriptive statistics in Chapter 2 of your text.

1.5 Editing and Printing Your Output

Once you have produced output in the Display Pane, you will likely want to edit and/or print it. For example, you may wish to add your name, date and a class assignment number. You may then want to print out the result to turn in to your instructor.

Before beginning the editing process, it will be helpful to understand the relationship between the items in the Outline Pane and various components of the Display Pane. Each element of the Outline Pane corresponds to a specific component of the Display Pane. To see how this works, click on **Log** in the Outline Pane. Notice that a box appears around the scripting commands at the top of the Display Pane indicating that Log corresponds to this portion of the Display Pane. Click a few other items in the Outline Pane to see how they correspond to portions of the Display Pane.

Begin editing by clicking the **Log** item in the Outline Pane. We wish to insert some text under the scripting language commands so click **Insert⇒New**

Figure 1.8: Edited output of basic statistics from x variable in text Table 2.6.

→ Cliff Blair 9/12/07 Biostatistics 501 Assignment 1					
Descriptives					
[DataSet1] C:\aamystuff\Books\Basic\aa_SPSS_manual\table 2.6 data.sav					
Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
x	11	3	5	4.00	.775
Valid N (listwise)	11				

Text. A box with cross hatching appears beneath the scripting commands in the Display Pane.⁴ Click inside the box. We want to bold your inserted text so click the bold button **B**. Now type your name followed by a space and then the date. Press **Enter** to move to the next line. (Your typed text may disappear but don't worry, it's still there.) Now type *Biostatistics 501 Assignment 1*. Click anywhere outside the box. (You may have to click twice to make the box disappear completely.) Your inserted text is there in all its bold faced glory. Let's clean up the output a bit more by getting rid of the scripting statements at the top of the Display Pane. To do this click the **Log** item in the Outline Pane. Now click **Edit⇒Delete**. The word **Log** disappears from the Outline Pane as does the scripting language statements from the Display Pane. Our edited output appears in Figure 1.8. You can delete other items in the same fashion. You can change the relative position of elements in the Display Pane by clicking and dragging them up or down in the Outline Pane. You can print or save your output by using the **File⇒Print...** and **File⇒Save As...** commands respectively. You can close the output file as we will not need it again.

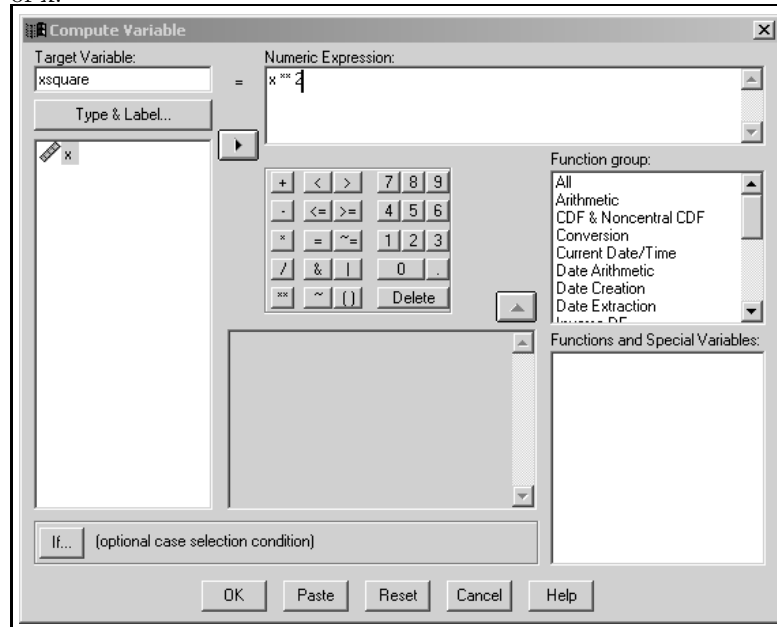
1.6 Creating New Variables



From time to time you may wish to use the variables in your data set to create new variables. This can be done directly from the Data Editor. To demonstrate, let's create the second column in Table 2.6 which, as you can see, is the square of the first column.

To begin, be sure that **table 2.6 data.sav** is loaded into the Data Editor. Then, from the Data Editor click **Transform⇒Compute Variable...**. The **Compute Variable...** window opens. In the **Target Variable:** window type *xsquare*. This will be the name of the new variable which will be the square of x. All variables in the dataset will be listed in the **Type & Label...** window. In

⁴If there is no cross hatching you may need to double click the box. The cross hatching means the box is in editing mode.

Figure 1.9: Compute Variable screen used to create the square of x .

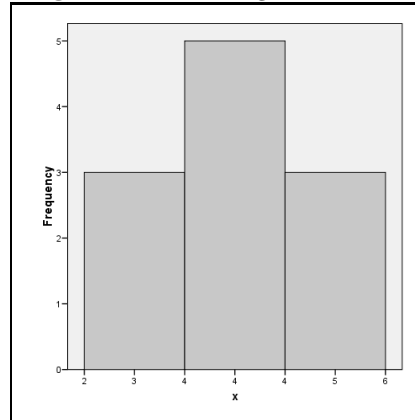


this case, only x is listed because x is the only variable in the dataset. Select (i.e. click on) x in the **Type & Label...** window then click the  button to move a copy of x into the **Numeric Expression:** window. Now click the  button on the keypad to move this expression into the **Numeric Expression:** window. In SPSS, as well as in some other languages, $**$ means “raise to the power of”. Now click 2 on the keypad to place this number in the numeric expression since we are raising x to the second power. Your screen should now appear as in Figure 1.9. Notice that we have simply written an algebraic expression that says $xsquare$ equals x raised to the second power.

Now click **OK**. The Data Editor shows a variable named $xsquare$ in the second column with associated observations that are the square of x . Click **File**⇒**Save** to save our work.

As a second example, let’s generate the values in the column headed $x - \bar{x}$ in Table 2.6. These values are termed “deviation scores” and will be explained in text Chapter 2. To calculate these deviation scores, we must subtract the average (i.e. mean) of x from each value of x . Unfortunately, SPSS does not provide a simple, intuitive means of calculating deviation scores “on the fly” (though it can be done). So we will take a simple approach based on the fact that we previously found the mean of x (see page 8 of this manual) to be 4.0.

With this knowledge we can write an expression to calculate the desired deviation scores. To do so, type *deviation* in the **Target Variable:** window and $x - 4$ in the **Numeric Expression:** window. Now click **OK** and close, without

Figure 1.10: Histogram of x.

saving, the Output Window. Notice that the deviation scores are now in the third column of the Data Editor. Save the new dataset.

1.7 Graphs

SPSS can be used to create a wide variety of graph types ranging from the simple to the complex. You will employ some of these capabilities in Chapter 2 of this manual. For the moment, we will provide only a simple example of one graph type by constructing a histogram of the x variable in Table 2.6.

Over time, SPSS has developed three different methods for constructing graphs. We will use the newest and easiest of these. To begin, be sure **table 2.6 data.sav** is the current worksheet in the Data Editor. Then click **Graphs⇒Chart Builder**.⁵ The Chart Builder window opens. In the **Choose from:** window click *Histogram*. Notice the graph depictions to the right of the **Choose from:** window. The first of these is a simple histogram. If you place the cursor over this depiction a short description will appear to confirm that this is the Simple Histogram.

Click and drag the Simple Histogram onto the “canvas” which is the large area above the Gallery. A histogram is now on the canvas but this is not *your* histogram but rather is an example histogram. Now drag the x variable from the **Variables:** window to the box labeled “X-Axis?” that extends along the x axis of the graph. (If you want to start this process over at any time simply click the **Reset** button to clear the canvas.) Now click **OK** and the graph depicted in Figure 1.10 appears.

⁵The other two methods for constructing graphs are termed **Interactive** and **Legacy Dialogs**.

Figure 1.11: Reconstruction of Table 2.6.

	x	xsquare	deviation	absolute deviation	squared deviation
1	3	9.00	-1.00	1.00	1.00
2	3	9.00	-1.00	1.00	1.00
3	3	9.00	-1.00	1.00	1.00
4	4	16.00	.00	.00	.00
5	4	16.00	.00	.00	.00
6	4	16.00	.00	.00	.00
7	4	16.00	.00	.00	.00
8	4	16.00	.00	.00	.00
9	5	25.00	1.00	1.00	1.00
10	5	25.00	1.00	1.00	1.00
11	5	25.00	1.00	1.00	1.00

Exercises

- 1.1 Use the **Compute Variable** feature of SPSS to construct the remaining variables in Table 2.6. The result should appear as in Figure 1.11. (Hint: You can place the absolute value function in the **Numeric Expression:** window by clicking *Arithmetic* in the **Function group:** window, then double-clicking the *Abs* function in the **Functions and Special Variables:** window.
- 1.2 Use the **Descriptives...** procedure to find the means of all the variables in Table 2.6. Don't generate any other statistics. Use **Descriptives...** only once. (Hint: You can move multiple variables into the **Variables(s):** window.
- 1.3 Label the output from Exercise 1.2 with your name and the current date. Remove all output above the line **Descriptive Statistics** in the printout. Print out the result.
- 1.4 Enter the data from text Table 7.1 on page 266 into a new worksheet. Save the worksheet with the name **table 7.1 data**.
- 1.5 Find the means of the three diet groups in Table 7.1.

Chapter 2

Descriptive Methods

2.1 Introduction

In this chapter you will learn to use SPSS to construct the graphs and calculate most of the statistics presented in text Chapter 2. You will also learn that in a few instances SPSS and the text do not agree on the calculated values of certain descriptive statistics because different definitions of these statistics are used by the software and text. You will also be afforded an opportunity to practice data entry and data transformations.

2.4 Distributions (page 15)

Task 2.1

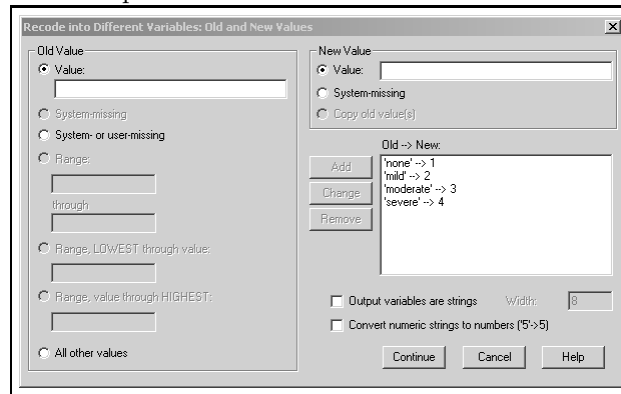
Use the data in Table 2.1 (**table 2.1 data.sav**) to construct frequency, relative frequency, cumulative frequency, and cumulative relative frequency distributions as shown in Table 2.2 on page 16.

Solution

Load the worksheet containing the Table 2.1 data by selecting **File⇒Open⇒Data...** and opening file **table 2.1 data.sav**. In Data View, note first that the Pain Level variable is expressed in text rather than numeric form. Because the cumulative frequency and cumulative relative frequency distributions require that the Pain Level variable be ordered, we must form a low to high ordering of this variable before we can construct these distributions. After all, SPSS has no way of knowing that “none” is a lesser pain level than is “mild” and that “mild” is a lesser pain level than is “moderate.” Our strategy will be to form a numeric variable from the text variable so SPSS can discern the correct ordering.

To do this, click **Transform⇒Recode into Different Variables....** The **Recode into Different Variables** window opens. Double-click the pain_level

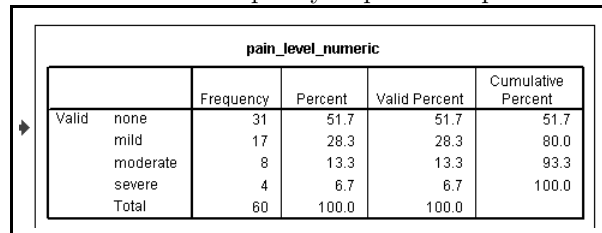
Figure 2.1: Screen appearance for assigning numeric values to pain level variable.



variable in the left most window to move it to the **Input Variable -> Output Variable:** window. After the move the window heading changes to **String Variable -> Output Variable:**. In the **Name:** portion of the **Output Variable** window type *pain_level_numeric* then click **Old and New Values:**. The **Recode into Different Variables: Old and New Values** window opens. Type *none* in the **Old Value** window and *1* in the **New Value** window. Now click the **Add** button. Notice that *'none' -> 1* appears in the **Old -> New** Window. Now type *mild* in the **Old Value** window and *2* in the **New Value** window and click **Add**. This adds *'mild' -> 2* to the **Old -> New** Window. Continue this process using the pairs *moderate*, *3* and *severe*, *4*. Your screen should now appear as in Figure 2.1. These statements inform SPSS as to how the new numeric variable is to be formed. That is, in each instance where **pain_level** is **none**, the new variable will take the value **1**, where **pain_level** is **mild**, **pain_level_numeric** will be **2** and so forth. Now click **Continue** which brings us back to the **Recode into Different Variables** screen. Click **Change** then **OK**. This brings us back to the Data Editor where we see that a new variable named **pain_level_numeric** has been added to the dataset.

Let's tidy up the Data Editor entries just a bit before continuing. To do this, click the **Variable View** tab at the bottom of the screen. As may be seen, the third row represents the newly created variable. Since we don't need decimal places for the integers 1, 2, 3 and 4, click the **Decimals** box and change the 2 to 0. If you go back to Data View you can see that the two decimal places have been removed from the *pain_level_numeric* scores. Back in Variable View change **Width** to 1 and **Columns** to 2. Back in Data View you can see that this removed much unneeded white space. These changes are not important but do tidy things up a bit. In Variable View click the **Measure** box and change the measure setting to *Ordinal*. After all, the variable we have created is clearly ordinal in quality. Let's also attach a label to the new variable which SPSS can use when displaying results of analyses, graphs etc. In Variable View click on

Figure 2.2: Frequency, relative frequency and cumulative relative frequency of perceived pain scores.



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	none	31	51.7	51.7	51.7
	mild	17	28.3	28.3	80.0
	moderate	8	13.3	13.3	93.3
	severe	4	6.7	6.7	100.0
	Total	60	100.0	100.0	

the **Label** box for the `pain_level_numeric` variable and type *Pain Level*. We will see the result of this labeling later.

If we were to form the desired distributions for the `pain_level_numeric` variable at this point, SPSS would give us distributions for pain levels 1, 2, 3 and 4 rather than none, mild, moderate and severe. To remedy this, let's make one other change. In Variable View, click in the **Value** box then click the button with three dots that appears therein. The **Value Labels** window opens. This will permit us to assign labels to each of the four values of `pain_level_numeric`. When providing output, SPSS will use these labels rather than the variable values 1 through 4. Type *1* in the **Value:** box and *none* in the **Label:** box. Click **Add**. Continue this process using the pairs 2 and mild, 3 and moderate, and 4 and severe. When you are finished click **OK**. Save the data sheet for later use by clicking **File**⇒**Save**.

We are at last ready to construct the desired distributions. To do so, from the Data Editor click **Analyze**⇒**Descriptive Statistics**⇒**Frequencies...**. The **Frequencies** window opens. Move the `pain_level_numeric` variable to the **Variable(s):** window. Be sure there is a check in the **Display frequency tables** box then click **OK**. The relevant portion of the output is shown in Figure 2.2.

Several points regarding this output should be noted. First, the categories are correctly labeled as none, mild, moderate and severe rather than 1, 2, 3 and 4. This is because of the labels we attached to the numeric categories. Second, what your text refers to as Relative Frequency and Cumulative Relative Frequency are labeled Percent and Cumulative Percent by SPSS. This is not unusual as many terms and styles of presentation are commonly used by both textbook authors and software developers. This may present some difficulties, but they tend to diminish as you gain more experience with different software packages. Third, with the exception that SPSS uses percents where the text uses proportions and the number of decimal places reported, the distributions provided by SPSS are the same as those in text Table 2.2. Fourth, SPSS orders the distributions in descending order while the text uses an ascending order. The style used by SPSS is the more common. The text authors prefer the ascending ordering but either is correct and provides the same information. Finally, SPSS does not provide the cumulative frequency distribution for the variable.

Task 2.2

Use the data in worksheet **table 2.3 data.sav** to construct the simple frequency distribution depicted in the text Table 2.3 on page 18.

Solution

Because the blood pressure data in **table 2.3 data.sav** are numeric as opposed to text, we need not go through the process of defining the variable order as was necessary with Task 2.1. Thus, we can construct the simple frequency distribution without any preliminaries.

After loading **table 2.3 data.sav** we select **Analyze⇒Descriptive Statistics⇒Frequencies...** When the **Frequencies** window opens move **BP** into the **Variables(s):** window and click **OK**. The relevant portion of the output is shown in Figure 2.3. Notice that no cumulative frequency distribution is provided.

Task 2.3

Use the data in worksheet **table 2.3 data.sav** to construct the grouped distributions depicted in text Table 2.4 on page 18.

Solution

Load the **table 2.3 data.sav** dataset into the Data Editor if it isn't already there. Before constructing the grouped distributions, we must first create a new variable that groups the blood pressure scores into the desired categories. This process is called *binning*. To do so, click **Transform⇒Visual Binning...** The **Visual Binning** window opens. Move the *BP* variable from the **Variables:** window to the **Variables to Bin:** window. Click **Continue**. Select (i.e. click on) the *BP* variable in the **Scanned Variable List:** window. Notice that SPSS shows you a histogram as well as the minimum and maximum values of our data. Let's name the new (binned) variable *BP12* by typing this in the **Binned Variable:** window. We can label the new variable by typing a label name in the **Label:** window. We'll use the label *BP in 12 Categories*.

We are now ready to create the categories. To do so, click the **Make Cutpoints...** button. The **Make Cutpoints** window opens. Type *89* in the **First Cutpoint Location:** window, *5* in the width window and *12* in the **Number of Cutpoints:** window. Your screen should appear as in Figure 2.4. Click **Apply**.

We will now make the appropriate labels for our categories by clicking the **Make Labels** button. Notice that SPSS has already created labels for us in the **Label** column. In order to conform with the text, change the first entry in this column from ≤ 89 to $85-89$. The remaining labels can be left as they are. Your screen should now appear as in Figure 2.5. You can see the characteristics of the new variable, including the value labels, in Variable View. Click **OK** then click **OK** again if SPSS warns you that one variable is to be created. From the Data Editor you can see that a new variable named *BP12* has been formed. The

Figure 2.3: Simple frequency distribution of BP scores as shown in Table 2.3.

BP				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 86	1	.7	.7	.7
90	1	.7	.7	1.4
92	2	1.4	1.4	2.8
93	1	.7	.7	3.5
95	3	2.1	2.1	5.6
96	2	1.4	1.4	6.9
97	2	1.4	1.4	8.3
98	2	1.4	1.4	9.7
99	1	.7	.7	10.4
100	4	2.8	2.8	13.2
102	1	.7	.7	13.9
103	1	.7	.7	14.6
105	2	1.4	1.4	16.0
106	1	.7	.7	16.7
107	2	1.4	1.4	18.1
109	1	.7	.7	18.8
110	3	2.1	2.1	20.8
111	3	2.1	2.1	22.9
114	2	1.4	1.4	24.3
115	6	4.2	4.2	28.5
116	3	2.1	2.1	30.6
117	1	.7	.7	31.3
118	2	1.4	1.4	32.6
119	2	1.4	1.4	34.0
120	3	2.1	2.1	36.1
121	1	.7	.7	36.8
122	3	2.1	2.1	38.9
123	2	1.4	1.4	40.3
124	4	2.8	2.8	43.1
125	4	2.8	2.8	45.8
126	7	4.9	4.9	50.7
127	3	2.1	2.1	52.8
128	3	2.1	2.1	54.9
129	3	2.1	2.1	56.9
130	5	3.5	3.5	60.4
131	3	2.1	2.1	62.5
132	4	2.8	2.8	65.3
133	8	5.6	5.6	70.8
134	5	3.5	3.5	74.3
135	8	5.6	5.6	79.9
136	3	2.1	2.1	81.9
137	11	7.6	7.6	89.6
138	3	2.1	2.1	91.7
139	6	4.2	4.2	95.8
140	4	2.8	2.8	98.6
143	2	1.4	1.4	100.0
Total	144	100.0	100.0	

Figure 2.4: Bin definitions used to form BP measures into 12 categories.

Make Cutpoints

☒ Equal Width Intervals

Intervals - fill in at least two fields

First Cutpoint Location: 89

Number of Cutpoints: 12

Width: 5

Last Cutpoint Location: 144

☐ Equal Percentiles Based on Scanned Cases

Intervals - fill in either field

Number of Cutpoints:

Width(%):

☐ Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases

☐ +/- 1 Std. Deviation

☐ +/- 2 Std. Deviation

☐ +/- 3 Std. Deviation

Apply will replace the current cutpoint definitions with this specification.
A final interval will include all remaining values: N cutpoints produce N+1 intervals.

Apply Cancel Help

Figure 2.5: Completed visual binning screen used to form BP measures into 12 categories.

Visual Binning

Scanned Variable List:

Variable
BP [BP]

Current Variable: BP

Binned Variable: BP12

BP in 12 Categories

Minimum: 86 Nonmissing Values Maximum: 143

Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value of 10, for example, defines an interval starting above the previous interval and ending at 10.

Grid	Value	Label
1	89	85-89
2	94	90 - 94
3	99	95 - 99
4	104	100 - 104
5	109	105 - 109
6	114	110 - 114
7	119	115 - 119
8	124	120 - 124
9	129	125 - 129
10	134	130 - 134
11	139	135 - 139
12	143	140 - 143

Cases Scanned: 144

Missing Values: 0

Copy Bins

From Another Variable...

To Other Variables...

Upper Endpoints

☒ Included (<=)

☐ Excluded (<)

Make Cutpoints...

Make Labels

☐ Reverse scale

OK Paste Reset Cancel Help

Figure 2.6: Grouped distributions of BP values in 12 categories.

BP in 12 Categories					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	85-89	1	.7	.7	.7
	90 - 94	4	2.8	2.8	3.5
	95 - 99	10	6.9	6.9	10.4
	100 - 104	6	4.2	4.2	14.6
	105 - 109	6	4.2	4.2	18.8
	110 - 114	8	5.6	5.6	24.3
	115 - 119	14	9.7	9.7	34.0
	120 - 124	13	9.0	9.0	43.1
	125 - 129	20	13.9	13.9	56.9
	130 - 134	25	17.4	17.4	74.3
	135 - 139	31	21.5	21.5	95.8
	140 - 144	6	4.2	4.2	100.0
	Total	144	100.0	100.0	

values of BP12 represent the 12 categories to which each value of BP belongs. Save your data by clicking **File⇒Save**. We will use it again later.

We are at last ready to form the specified distributions. Click **Analyze⇒Descriptive Statistics⇒Frequencies....** Move the “BP in 12 Categories” variable to the **Variable(s):** window and click **OK**. The relevant portion of the output is shown in Figure 2.6. We again note that with the exception that SPSS does not provide a cumulative frequency distribution, uses percents where the text uses proportions and the number of decimal places reported, the distributions provided by SPSS are the same as those in text.

2.5 Graphs (page 19)

Task 2.4

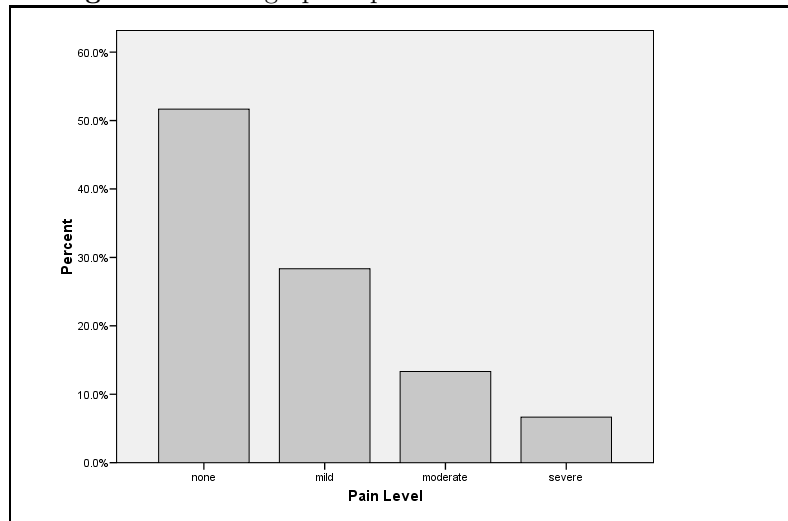
Use the data in Table 2.1 (**table 2.1 data.sav**) to construct the relative frequency bar graph depicted in text Figure 2.1 on page 20.

Solution

Open **table 2.1 data.sav** into the Data Editor. Remember that we modified this dataset in Task 2.1. Select **Graphs⇒Chart Builder....**¹ The Chart Builder window opens. If a graph has been previously constructed by Chart Builder it will appear in the “canvas” which is the large area at the top of the Gallery. If this is the case click the **Reset** to clear the canvas. Reset can be used at any time during the graph building process to start the process over.

Click on **Gallery** to be sure it is active. Click on *Bar* in the **Choose from:** list. Notice the graph depictions to the right of the **Choose from:** window. The first of these is a simple bar graph. If you place the cursor over this depiction a short description will appear to confirm that this is the Simple

¹SPSS uses the terms chart and graph synonymously.

Figure 2.7: Bar graph of pain level scores.

Bar graph. Click and drag the Simple Bar graph onto the canvas area. A bar graph appears on the canvas but this is not *your* bar graph but rather is an example bar graph. When you dragged the Simple Bar depiction onto the canvas the **Element Properties** window should have opened. If it didn't, click **Element Properties....** In the **Edit Properties of:** window of the **Element Properties** panel, click *Bar1*. Then in the **Statistic:** window of the same panel choose *Percentage ()* then click **Apply** at the bottom of the panel. This mandates that the bars will represent percentages of observations rather than counts or some other statistic.

Now drag the Pain Level [pain_level_numeric] variable from the **Variables:** window to the box labeled "X-Axis?" that extends along the x axis of the graph. (If you want to start this process over at any time simply click the **Reset** button to clear the canvas.) Now click **OK** and the graph depicted in Figure 2.7 appears. Notice that the pain levels are ordered and bars are properly labeled. How do you think SPSS knew how to do this? (Hint: go back and look at Task 2.1.)

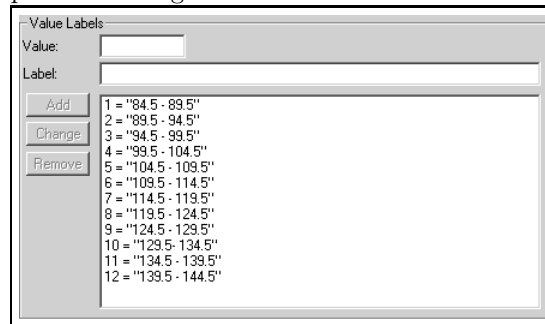
Task 2.5

Use the data in Table 2.3 (**table 2.3 data.sav**) to construct the relative frequency histogram depicted in text Figure 2.2 on page 21.

Solution

Be sure **table 2.3 data.sav** is the active worksheet. We could use the BP12 variable we created earlier in connection with Task 2.3 as the basis of the grouped histogram but the labels we associated with the values of that variable do not

Figure 2.8: Labels for the 12 group blood pressure histogram.



reflect the upper and lower limits of the intervals as is done in Figure 2.2. Our solution will be to copy BP12 to a new variable and edit the labels associated with the new variable to reflect the values depicted in Figure 2.2.

To this end, click the **Data View** tab if you aren't already in Data View. Click the **BP12** designation at the top of the second data column. Notice that the entire column darkens to indicate that this column has been selected. Now click **Edit⇒Copy** to copy this column. Click the **Var** designation at the top of the next (empty) column. Now this column darkens. Click **Edit⇒Paste**. A copy of the BP12 data has been pasted into the third column. SPSS has changed the name of the copied data. After all, we can't have two variables with the same name.

Now click the **Variable View** tab. Click the **Name** box in the third row and type *BP12_hist*. Check the **Measure** box for BP12_hist to be sure it is set to "ordinal." SPSS takes seriously the level of measurement of variables, especially when constructing graphs. We must now edit the labels of BP12_hist to have them better conform to the labeling of the x axis in Figure 2.2.

Click the **Values** box in row 3 then click the small icon with three dots that appears. The **Value Labels** window opens showing the labels we created in conjunction with Task 2.3. Click the label 1="85 - 89." In the **Label:** box edit 85 - 89 to read 84.5 - 89.5 then click **Change**. Now click the label 2="90 - 94" and edit this to read 2="89.5 - 94.5" and again, click **Change**. Continue editing until the labels appear as in Figure 2.8. Notice that I removed the 13th category label by selecting the category and clicking the **Remove** button. Click **OK** to make the changes and close the **Value Labels** window. It would be a good idea to save your dataset at this point. We are now ready to construct the histogram.

To this end click **Graphs⇒Chart Builder....** The **Chart Builder** window opens. Click the **Gallery** button to be sure this option is selected. If a chart is present from a previous session click **Reset** to clear the chart area. Select *Histogram* in the **Choose from:** window. Four graph depictions appear to the right of the **Choose from:** window. The first of these, as can be verified

by placing the cursor over the depiction, is the **Simple Histogram** graph. Click and drag this chart onto the large canvas area. A histogram appears but this is not *your* histogram but is rather an example to be modified into your histogram. At the same time, the **Element Properties** window opens. Now click and drag the BP12_hist variable into the horizontally positioned box that is labeled “X-Axis?” below the example histogram. In the **Edit Properties of:** window select *Bar1* if it isn’t already highlighted. Then select *Percentage()* in the **Statistic** window. We are using percentage because proportion is not available in SPSS. Click the **Apply** button at the bottom of the **Element Properties** window to apply these changes to your graph. Now click **OK** at the bottom of the **Chart Builder** window to construct your graph.

You may have to scroll down to see your graph. As expected, the bars represent percents rather than proportions and the labeling of the x axis is not exactly as depicted in Figure 2.2 but this is about the best we can do in this regard without considerable effort. But there is still something wrong. The bars are spaced as would the bars of a bar graph rather than being contiguous as would be the case for a histogram. We will use the Chart Editor to correct this situation. Double-click anywhere in the graph. The **Chart Editor** opens. Now double-click on one of the bars. The **Properties** window opens. Click the **Bar Options** tab. Two sliders labeled **Bars** and **Clusters** are revealed. Slice the **Bars** slider all the way to the right (The % window will go to 100) and click **Apply** then **Close**. Close the Chart Editor by clicking the close button at the top right corner of the screen or selecting **File⇒Close**. Your graph appears as shown in Figure 2.9.

Task 2.6

Use the data in Table 2.3 (**table 2.3 data.sav**) to construct the relative frequency polygon depicted in text Figure 2.4 on page 22.

Solution

We will use a slightly modified form of the labels we created in connection with Task 2.5 to label this graph. We could edit the labels associated with the variable BP12_hist for this purpose but instead will create a new variable so we can preserve our previous work.

Be sure the **table 2.3 data.sav** data is loaded into the Data Editor. From Data View, select the column heading BP12_hist as shown in Figure 2.10. Click **Edit⇒Copy** then click the **Var** heading of the adjoining column and click **Edit⇒Paste**. This pastes a copy of BP12_hist into the previously empty column. Notice, however, that SPSS has changed the name of the new variable to avoid confusion with BP12_hist. In Variable View click the **Name** box for the new variable and type *BP12_poly*. We now need to edit the labels associated with BP12_poly for reasons to be noted below.

Click the **Values** box for BP12_poly then click the icon with three dots that appears. The **Value Labels** window opens revealing the labels we previously created for BP12_hist. Type 0 in the **Value:** window and 79.5-84.5 in the

Figure 2.9: SPSS rendering of the histogram depicted in text Figure 2.2.

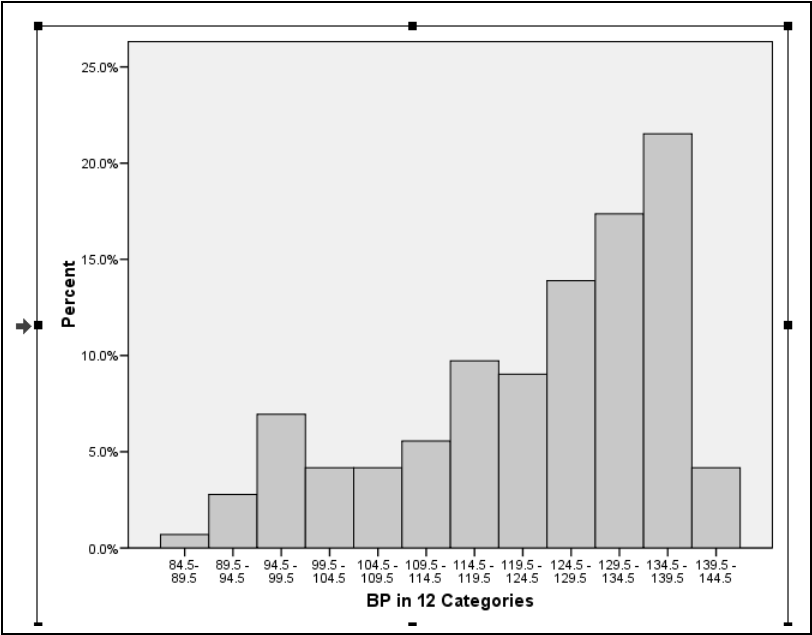
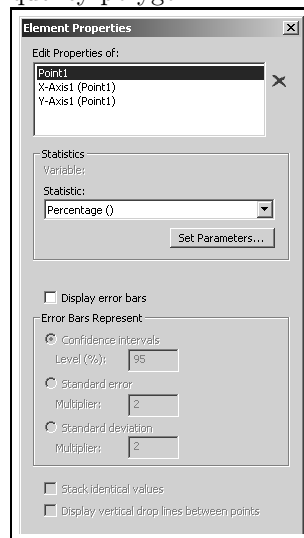


Figure 2.10: Selecting the BP12_hist variable in Data View.

	BP	BP12	BP12	var
1	140	12	12	
2	128	9	9	
3	137	11	11	
4	143	12	12	
5	136	11	11	
6	129	9	9	
7	137	11	11	
8	143	12	12	
9	120	8	8	
10	119	7	7	
11	135	11	11	
12	118	7	7	
13	138	11	11	
14	109	5	5	
15	100	4	4	
16	129	9	9	
17	135	11	11	
18	100	4	4	
19	120	8	8	
20	114	6	6	
21	114	6	6	
22	135	11	11	

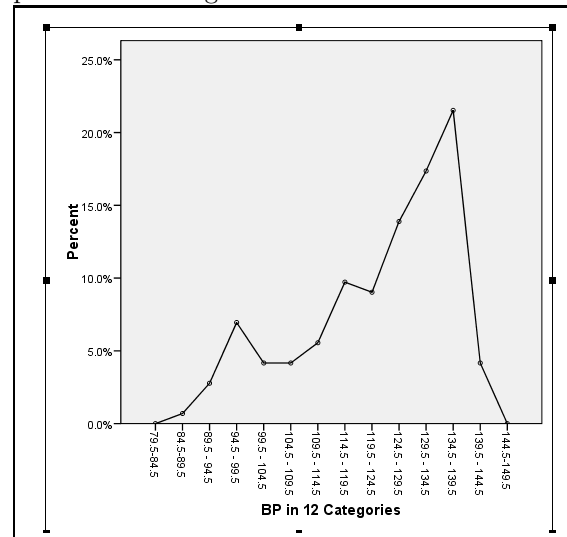
Figure 2.11: Element Properties screen used in the construction of the 12 category BP relative frequency polygon.



Label: window then click **Add**. The new label is added to the list of labels. We must add one more label to complete our task. Type *13* in the **Value:** window and *144.5-149.5* in the **Label:** window then click **Add**. Notice that there are no observations in the intervals 79.5-84.5 or 144.5-149.5. This means that SPSS will plot these points on the polygon as zero which will bring the line down to the x axis at either end of the graph as shown in text Figure 2.4. Click **OK** to preserve the label changes and save the data by clicking **File**⇒**Save**.

We are now ready to construct the polygon. Our strategy will be to create a scatter plot then instruct SPSS to connect the points with lines. To begin click **Graphs**⇒**Chart Builder...** The Chart Builder window opens. Be sure **Gallery** is selected. If a graph from a previous task is occupying the canvas, click **Reset** to clear the space. Select *Scatter/Dot* then click and drag the first chart on the top row onto the canvas. If you are not sure which chart to drag just place your cursor over the charts and look for the one designated **Simple Scatter**. An example scatter plot appears on the Gallery canvas. Drag the BP12_poly variable into the horizontally positioned box that is labeled “X-Axis?” below the example scatter plot. On the **Element Properties** screen select **Point1** in the **Edit Properties of:** window and *Percentage (?)* in the **Statistic:** window. Your **Edit Properties** screen should appear as in Figure 2.11. Click **Apply** to activate the changes then click **OK** to construct the scatter plot.

Figure 2.12: SPSS rendering of the polygon depicted in text Figure 2.4.



We will need the Chart Editor to complete our task. Double-click on the graph to open the Chart Editor. The dots that make up the scatter plot look a little big so let's begin by reducing their size slightly. To do this click on one of the dots to select it. Then click **Edit**⇒**Properties**. The **Properties** window opens. Click the **Marker** tab and change the value in the **Size** window to 4. Click **Apply** to make this change effective. Click **Close** to close the **Properties** window. Click anywhere on the graph other than on a dot to de-select the dots and see the size reduction. Now click **Elements**⇒**Interpolation Line**. The dots are now connected. Close the Chart Editor by clicking **File**⇒**Close**. The relative frequency polygon appears as in Figure 2.12. Again, the graph depicts percents rather than proportions and the labeling of the x axis may not be exactly as desired but the graph is quite acceptable and about as good as we can do without considerably more effort.

Task 2.7

Use the data in Table 2.3 (**table 2.3 data.sav**) to construct the cumulative relative frequency polygon depicted in text Figure 2.6.

Solution

We will use the same general approach for completing this task as we used for Task 2.6. That is, we will copy the grouped blood pressure scores into a new location in the Data Editor, give this variable a new name, edit the labels associated with the new variable and use Chart Builder to create a scatter plot with connecting lines.

To begin, copy the BP12_poly variable into the first empty column in the Data Editor. Name the new variable BP12_cum_poly. (See Task 2.6 if you don't remember how these things are done.) In Variable View, click the **Values** box for BP12_cum_poly then click the icon with three dots that appears. The **Value Labels** window opens revealing the labels we previously created for BP12_poly. Recall that the label associated with the score of 13 was created to bring the upper end of the relative frequency polygon back to the x axis. We don't need to do this for the cumulative relative frequency polygon so select this label and click **Remove** then **OK**. Save your dataset.

We are now ready to construct our graph. To begin click **Graphs⇒Chart Builder...** The Chart Builder window opens. Be sure **Gallery** is selected. If a graph from a previous task is occupying the canvas, click **Reset** to clear the space. Select *Scatter/Dot* then click and drag the first chart on the top row onto the canvas. If you are not sure which chart to drag place your cursor over the charts and look for the one designated **Simple Scatter**. An example scatter plot appears on the Gallery canvas. Drag the BP12_cum_poly variable into the horizontally positioned box that is labeled "X-Axis?" below the example scatter plot. On the **Element Properties** screen select **Point1** in the **Edit Properties of:** window and *Cumulative Percentage* in the **Statistic:** window. Click **Apply** to activate the changes then click **OK** to construct the scatter plot.

We will need the Chart Editor to complete our task. Double-click on the graph to open the Chart Editor. The dots that make up the scatter plot look a little big so let's begin by reducing their size slightly. To do this click on one of the dots to select it. Then click **Edit⇒Properties**. The **Properties** window opens. Click the **Marker** tab and change the value in the **Size** window to 4. Click **Apply** to make this change effective. Click **Close** to close the **Properties** window. Click anywhere on the graph other than on a dot to de-select the dots and see the size reduction. Now click **Elements⇒Interpolation Line**. The dots are now connected. Close the Chart Editor by clicking **File⇒Close**. The cumulative relative frequency polygon appears as in Figure 2.13. Again, the graph depicts percents rather than proportions and the labeling of the x axis may not be exactly as desired but the graph is quite acceptable and about as good as we can do without considerably more effort.

Task 2.8

Construct a stem-and-leaf-display for the blood pressure data in text Table 2.3 on page 18.

Solution


Be sure that **table 2.3 data.sav** is the current worksheet. Click **Analyze⇒Descriptive Statistics⇒Explore...** The **Explore** window opens. Select the BP variable and click the  button to move it into the **Dependent List:** window. Darken the **Plots** button under **Display**. Your screen should appear as in Figure 2.14. Now click the **Plots...** button. The **Explore:Plots** window opens. Be sure the

Figure 2.13: SPSS rendering of the cumulative relative frequency polygon depicted in text Figure 2.6.

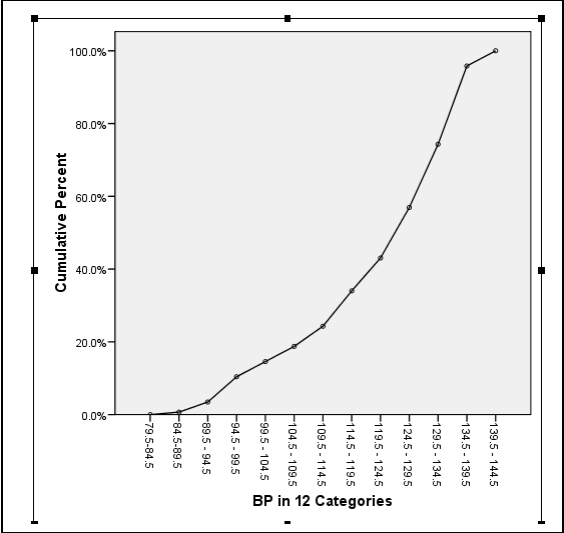


Figure 2.14: Screen appearance of Explore window when creating a stem-and-leaf-display.

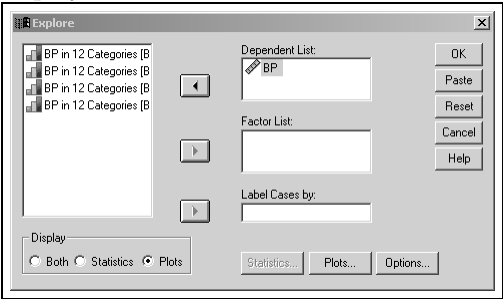
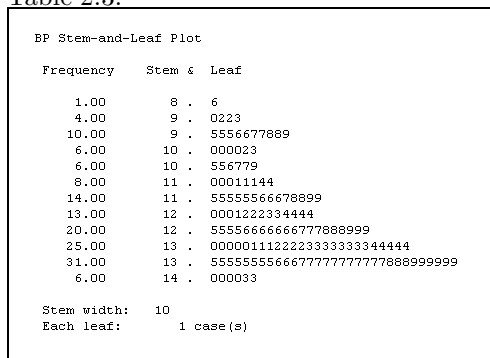


Figure 2.15: SPSS rendering of a stem-and-leaf display for the BP variable in text Table 2.3.



Stem and leaf box is checked and uncheck the **Histogram** box if it is checked. Click **Continue** then **OK**. The stem-and-leaf-display appears as shown in Figure 2.15

Note that SPSS divides leaves in half so that digits 0 through 4 are placed on one line and 5 through 9 are placed on a second line. Also, the numbers on the far left give the leaf count for each line.

2.6 Numerical Methods (page 23)

Task 2.9

In text Example 2.4 on page 27 you are asked to find the median of the numbers 3, 2, 0, 2, 1, 1, 2, and 2. Two different methods are used that produce two different answers, namely, 2.00 and 1.75. It is argued that the better answer is provided by the method that produces the median value of 1.75. What answer is provided by SPSS?

Solution

We begin by entering the numbers in a new worksheet. To get a blank worksheet click **File**⇒**New**⇒**Data**. In Data View, enter the 8 observations in the first column. Name the new variable x . Just to practice previously learned skills, get rid of the decimal portion of the integers you just entered. Your Data View screen should appear as in Figure 2.16.

Now click **Analyze**⇒**Descriptive Statistics**⇒**Explore....** The Explore window opens. Move x to the **Dependent List:** window. Choose the *Statistics* option under **Display** then click the **Statistics...** button. The **Explore: Statistics** window opens. Be sure the **Descriptives** box is checked then click **Continue** and **OK**. The relevant portion of the output is shown in Figure 2.17.

Figure 2.16: Data View for data used to calculate the median.

	x	var	var
1	3		
2	2		
3	0		
4	2		
5	1		
6	1		
7	2		
8	2		
9			

Figure 2.17: Calculation of the median of a group of scores.

Descriptives				
			Statistic	Std. Error
x	Mean		1.63	.324
	95% Confidence Interval for Mean	Lower Bound	.86	
		Upper Bound	2.39	
	5% Trimmed Mean		1.64	
	Median		2.00	
	Variance		.839	
	Std. Deviation		.916	
	Minimum		0	
	Maximum		3	
	Range		3	
	Interquartile Range		1	
	Skewness		-.488	.752
	Kurtosis		.421	1.481

As may be seen, the median value calculated by SPSS is 2.0. This is the value provided by most statistics software packages.

Task 2.10

Find the mean, median, and standard deviation of the data in the table associated with Example 2.7 on page 30. Is the median reported by SPSS the same as that reported in the text? Why do you think this is the case?

Solution

We will have to enter this data into a worksheet since none is provided. This could be a daunting task as there are 145 observations in this table. Fortunately, the data are provided in the form of a simple frequency distribution which is a contingency foreseen by SPSS. In a new worksheet, enter the scores in the first column with their associated frequencies in the second. Name the scores *x* and the frequencies *freq*. When entry is complete your Data Editor screen should appear as in Figure 2.18. We must now inform SPSS of this data arrangement.

To do so, click **Data⇒Weight Cases...** The **Weight Cases** window opens. Select the *Weight cases by* option then move the *freq* variable into

Figure 2.18: Data from example 2.7.

	x	freq	var	var	var	var	var
1	2.4	3					
2	2.3	40					
3	2.2	36					
4	2.1	21					
5	2.0	18					
6	1.9	8					
7	1.8	14					
8	1.7	5					
9							

Figure 2.19: Descriptive statistics for data from example 2.7.

Descriptives			Statistic	Std. Error
x	Mean		2.120	.0151
	95% Confidence Interval for Mean	Lower Bound	2.090	
		Upper Bound	2.150	
	5% Trimmed Mean		2.129	
	Median		2.200	
	Variance		.033	
	Std. Deviation		.1813	
	Minimum		1.7	
	Maximum		2.4	
	Range		.7	
	Interquartile Range		.3	
	Skewness		-.742	.201
	Kurtosis		-.470	.400

the **Frequency Variable:** window. Click **OK**. SPSS now understands that there are multiple entries of the x variable with the number of multiples being indicated by entries in the freq variable.

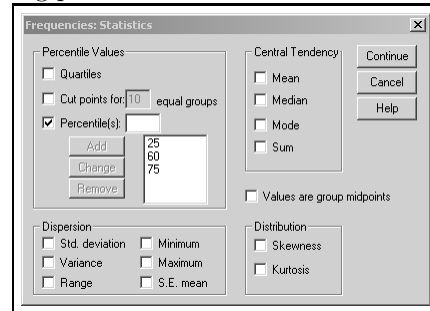
From this point we can proceed as in Task 2.9. That is, click **Analyze⇒Descriptive Statistics⇒Explore....** The Explore window opens. Move x to the **Dependent List:** window. Choose the *Statistics* option under **Display** then click the **Statistics...** button. The **Explore: Statistics** window opens. Be sure the **Descriptives** box is checked then click **Continue** and **OK**. The relevant portion of the output is shown in Figure 2.19.

As may be seen, the median reported here is 2.2 which contrasts with the value of 2.17 reported in the text as the answer for Example 2.7. Thus, it appears that SPSS calculates the median differently than does the text. The mean and standard deviation are given respectively as 2.120 and .1813.

Task 2.11

Find the variance of the x variable in Table 2.6 (**Table 2.6 data.sav**). Is the result provided by SPSS the same as that given in the text for example 2.15 on page 36?

Figure 2.20: Frequencies: Statistics panel configuration for calculating percentiles.



Solution

We could use the **Explore** command as we did in Tasks 2.10 and 2.11 but to gain experience with different procedures we will use a different command. After loading worksheet **Table 2.6 data.sav**,² select **Analyze⇒Descriptive Statistics⇒Descriptives...** The **Descriptives** window opens. Move the x variable into the **Variable(s):** window. Click **Options...** The **Descriptives: Options** window opens revealing various statistics that can be provided by this command. Place a check in the **Variance** box and uncheck the boxes associated with other statistics. Leave the selection under **Display Order** unchanged. Click **Continue** then **OK**. The output shows the variance to be .60 which is the value calculated in the text.

Task 2.12

Find the twenty-fifth, sixtieth, and seventy-fifth percentiles of the blood pressure data in worksheet **table 2.3 data.sav**. Do the answers provided by SPSS match those given in the solution to text Example 2.17 on page 38?

Solution

SPSS provides a simple means for finding percentiles through the **Frequencies** procedure. Load **table 2.3 data.sav** into the Data Editor. To find the required percentiles click **Analyze⇒Descriptive Statistics⇒Frequencies...** The **Frequencies** window opens. Move the BP variable into the **Variable(s):** window. Click **Statistics...** to open the **Frequencies: Statistics** panel. Put a check in the **Percentiles(s):** box and type 25 in the **Percentiles(s):** window. Click **Add** to move this value into the list of percentiles to be calculated. Repeat the process to place the values 60 and 75 in this list. Your screen should appear as in Figure 2.20.

Now click **Continue** then **OK** to produce the output seen in Figure 2.21.

²You created this dataset as outlined on page 5 of this manual.

Figure 2.21: Output from calculation of percentiles by SPSS.

Statistics		
BP		
N	Valid	144
	Missing	0
Percentiles	25	115.00
	60	130.00
	75	135.00

As may be seen, SPSS calculates the twenty-fifth, sixtieth, and seventy-fifth percentiles of the blood pressure measures to be 115, 130 and 135 respectively. This contrasts with the values of 114.67, 130.38 and 134.63 calculated in Example 2.17 on page 38 of the text. The discrepancy appears to arise because SPSS is using a score based definition of percentiles while the calculations in the text are scale based. The method used by SPSS appears to be the more common but the authors of your text prefer the scale based method.

Task 2.13

Use SPSS to perform the analyses outlined in Example 2.21 on page 43.

Solution

Open a new worksheet. Type the numbers 1, 3, 3, 9 in the first column. Name this variable x. Now click **Analyze**⇒**Descriptive Statistics**⇒**Descriptives...**. Move the x variable into the **Variables(s):** window. Place a check in the **Save standardized values as variables** box and click **OK**. Close the output window and return to the Data Editor. The z score for each of the x values is in the second column. SPSS has named this variable Z_X .

Notice that these z scores agree with those given in the text except that SPSS gives the z score for nine as 1.44338 while the text gives 1.445. The discrepancy occurs because the text uses 3.46 as the standard deviation while SPSS carries this calculation out to more decimal places. Now use the **Descriptives...** command to find the mean and standard deviation of Z_X . Review the solution for Task 2.11 if you don't remember how to use this command. As expected, the mean is given as zero and the standard deviation as one.

Task 2.14

Use SPSS to conduct the analyses alluded to in Example 2.23 on page 44 of the text.

Solution

Open the **table 2.8 data.sav** worksheet. The data for distributions A, B and C are in columns with the like letter. Click **Analyze**⇒**Descriptive Statistics**⇒**Descriptives...**

Move the A, B and C variables into the **Variable(s):** window. Click **Options...** to open the **Descriptives: Options** panel. Place a check in the **Skewness** box and uncheck any boxes related to other statistics. Click **Continue** then **OK**. The SPSS output gives skew for Distributions A, B and C as .000, $-.772$ and $.772$ respectively. The solution for Example 2.23 gives skew of .000, $-.648$, and $.648$. Because there are different definitions of skew, it appears that SPSS uses a different form than does the text.

Task 2.15

Use SPSS to perform the analyses outlined in Example 2.24 on page 46 of the text.

Solution

Open the **table 2.9 data** worksheet. Follow the same steps outlined in the solution to Task 2.14 with **Kurtosis** being checked rather than **Skewness** in the **Descriptive: Options** window. The values obtained for distributions A and B respectively are 5.50 and -1.65 which differs markedly from the values of 5.04 and 1.26 given in the text. This again occurs because SPSS uses a different definition of kurtosis than that used in the text.

Exercises

- 2.1 Use the data in worksheet **table 2.3 data.sav** to construct the grouped distributions depicted in the text Table 2.5.
- 2.2 Construct the histogram depicted in text Figure 2.3.
- 2.3 Construct a relative frequency polygon similar to that depicted in Figure 2.4 but with eight instead of 12 categories.
- 2.4 Construct a cumulative relative frequency polygon similar to that depicted in Figure 2.6 but with eight instead of 12 categories.
- 2.5 Use SPSS to find the mean and median of the following data.

<u>Score</u>	<u>frequency</u>
9	134
8	442
7	391
6	777

- 2.6 Calculate z scores for the BP variable in **table 2.3 data.sav**. Find the mean, standard deviation and variance of the resultant z scores.
- 2.7 Use SPSS to find the skew of the BP data in **table 2.3 data.sav**.

Chapter 3

Probability

3.1 Introduction

In this chapter you will learn to use SPSS to form contingency tables as well as to find areas under the normal curve.

3.3 Contingency Tables (page 52)

Task 3.1

Use worksheet **table 3.1 data.sav** to construct text Table 3.1 located on page 54.

Solution

Open the **table 3.1 data.sav** data into a fresh worksheet. Notice that the first column of **table 3.1 data.sav** characterizes each subject as being a smoker (S) or a non-smoker (SN) while the second column characterizes each subject as having some disease (D) or not having the disease (DN). Also take a moment to use Variable View to study how these two variables are characterized. The labeling of the two variables will permit us to improve on the rather terse form of text Table 3.1.

To form the table click **Analyze⇒Tables⇒Basic Tables....** The **Basic Tables** panel opens. Move the smoke variable into the **Down:** Window and the disease variable into the **Across** window. Now click the **Totals...** button to open the **Basic Tables: Totals** panel. Place a check in the **Table-margin totals (ignoring missing group data)** box. This causes SPSS to output the table margin totals. Now click **Continue** then **OK**. The relevant portion of the output is shown in Figure 3.1. Notice that the variables labels significantly improved the appearance of the table as compared to the austere form of text Table 3.1.

Figure 3.1: SPSS rendering of text Table 3.1.

		Disease Status		Table Total
		Disease	Disease Free	
Smoking Status	Smoker	9	3	12
	Non-Smoker	2	6	8
Table Total		11	9	20

Figure 3.2: Input to function for finding the area below 220 of a normal curve with mean 250 and standard deviation 25.

	point	mean	sd	var	var
1	220	250	25		
2					
3					

3.4 The Normal Curve (page 62)

Task 3.2

Use SPSS to find the area of the normal curve indicated in Example 3.9 on page 64.

Solution

SPSS uses functions to find the areas under various types of curves. Generally speaking, the user must provide specified parameters to the function which then returns the desired area. The function we will use to find the area under a normal curve requires that we specify (1) the point on the curve whose associated area is sought, (2) the mean of the distribution and (3) the standard deviation of the distribution. We begin by entering these values into a fresh worksheet.

Open a new worksheet. Type *220* in the first cell of the first row, then type *250* and *25* in the second and third cells of the same row. Name the first variable **point**, the second **mean** and the third **sd**. Your worksheet should appear as in Figure 3.2. Notice that we specified zero decimal places for these input values. We are now ready to invoke the normal curve area function.

To do so, click **Transform**⇒**Compute Variable....** The **Compute Variable** window opens. Type *area* in the **Target Variable:** window. Now find and select *CDF & Noncentral CDF* in the **Function group:** window. Double-click *Cdf. Normal* in the **Functions and Special Variables:** window. This places the expression *CDF.NORMAL(?,?,?)* in the **Numeric Expression:** window. Edit this expression replacing the question marks so that it reads *CDF.NORMAL(point,mean,sd)* then click **OK**. Returning to the Data Editor note that a new variable titled **area** has been added. This variable has an associated value of .12. If, in Variable View, you change the number in the

Decimals box to 4 then return to Variable View, you will find the value of the area variable to be .1151 which is the value given in the text.

Exercises

- 3.1 Construct a worksheet to represent the data in the accompanying contingency table. Then use the worksheet to construct the contingency table.

	B	\overline{B}	
A	1	3	4
\overline{A}	3	3	6
	4	6	

- 3.2 Use SPSS in lieu of the normal curve table (text Appendix A) to find the areas of the normal curve indicated in Examples 3.10, 3.11 and 3.12.

Chapter 4

Introduction to Inference and One Sample Methods

4.1 Introduction

In this chapter you will use SPSS to find binomial probabilities as well as probabilities associated with the normal curve. You will also learn to perform various hypothesis tests, including equivalence tests, and form certain confidence intervals. In addition, you will perform power calculations and compute required sample sizes to attain specified levels of power.

4.2 Sampling Distributions (page 75)

Task 4.1

Generate the binomial probabilities in Table 4.1 on page 83.

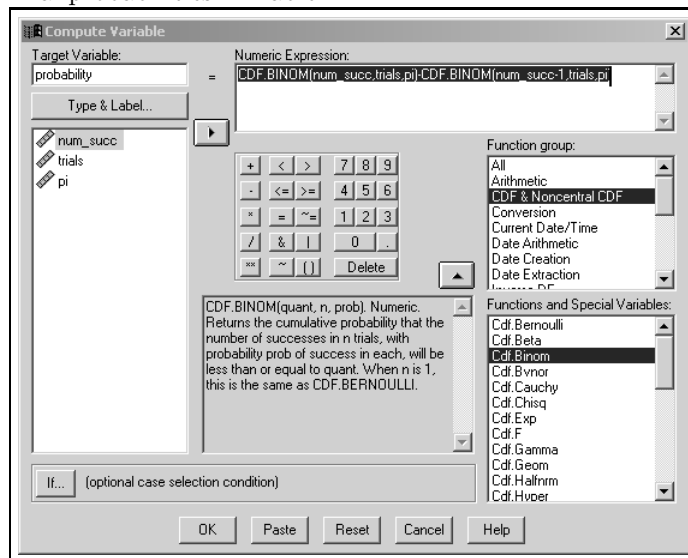
Solution

SPSS provides a function that calculates the probability of obtaining a given number of successes or a lesser number of successes for a binomial random variable. We will use this function to perform this task.

Generally speaking, the user must provide specified parameters to the function which then returns the desired probability. The function we will use to find the specified binomial probabilities requires that we specify (1) the number of successes for which the probability is to be computed, (2) the number of binomial trials and (3) the parameter π as described in the text. We begin by entering these values into a fresh worksheet.

Open a new worksheet. Type the integers 0 through 5 in the first column. These are the numbers of successes for which probabilities are to be calculated. Name this variable **num_succ**. Enter the integer 5 in the first six rows of the

Figure 4.1: Compute Variable screen for calculating binomial probabilities in Table 4.1.



second column. Name this variable **trials**. Finally, enter the decimal .1 in the first six rows of the third column. Name this variable **pi**.

To invoke the needed function click **Transform**⇒**Compute Variable...**. The **Compute Variable** window opens. Type *probability* in the **Target Variable:** window. Now find and select *CDF & Noncentral CDF* in the **Function group:** window. Double-click *Cdf. Binom* in the **Functions and Special Variables:** window. This places the expression *CDF.BINOM(?,?,?)* in the **Numeric Expression:** window. Edit this expression replacing the question marks so that it reads *CDF.BINOM(num_succ, trials, pi)*. Remember, this function returns the probability that the number of successes is less than or equal to **num_succ**. But our task is to find the probability that the number of successes will be equal to **num_succ**. To obtain this value we must subtract the probability that the number of successes is less than **num_succ** from the value returned by *CDF.BINOM(num_succ, trials, pi)*. To do this edit the expression in the **Numeric Expression:** window to read *CDF.BINOM(num_succ, trials, pi) - CDF.BINOM(num_succ - 1, trials, pi)*. Note that this expression subtracts the probability that the number of successes is less than **num_succ** from the probability that the number of successes is less than or equal to **num_succ**. Your **Compute Variable** window should now appear as in Figure 4.1.

Now click **OK**. Returning to the Data Editor note that a new variable titled **probability** has been added. In Variable View, change the number in the **Decimals** box to 5 then return to Variable View. The values of the probability variable are the same as those given in text Table 4.1.

Figure 4.2: Function parameters for finding the normal curve area below a Z score of -5.33 .

	Z	mean	sd	var
1	-5.33	0	1	
2				

Task 4.2

The Z value calculated in connection with Example 4.7 on page 86 of your text was -5.33 . Use SPSS to find the associated probability.

Solution

As with Task 3.2 on page 36 of this manual, we will use the function $CDF.NORMAL(?, ?, ?)$ to find the specified area of the normal curve. This problem differs slightly from that given in Task 3.2 in that the point on the normal curve is expressed as a Z score rather than a raw score. This means reference will be to a standard normal curve which has mean zero and standard deviation one.

Open a new worksheet. In the first row enter the values -5.33 , 0 and 1 in the first three columns. Name the first variable Z , the second mean and the third sd. Your screen should appear as in Figure 4.2.

Click **Transform** \Rightarrow **Compute Variable...** The **Compute Variable** window opens. Type *probability* in the **Target Variable:** window. Now find and select *CDF & Noncentral CDF* in the **Function group:** window. Double-click *Cdf. Normal* in the **Functions and Special Variables:** window. This places the expression $CDF.NORMAL(?, ?, ?)$ in the **Numeric Expression:** window. Edit this expression replacing the question marks so that it reads $CDF.NORMAL(Z, mean, sd)$ then click **OK**. Returning to the Data Editor note that a new variable titled **probability** has been added. This variable has an associated value of .00. In Variable View, change the number in the **Width** box to 10 and the number in the **Decimals** box to 8 for the probability variable then return to Variable View. As may be seen, the value of the probability variable is given as .00000005 which is the sought probability.

4.3 Hypothesis Testing (page 86)

Task 4.3

Use SPSS to perform the one mean Z test outlined in Example 4.9 on text page 92.

Solution

SPSS does not provide a straightforward means for conducting a one mean Z

Figure 4.3: Data Editor after performing calculations for Task 4.3.

	xbar	mu	sigma	n	alpha	obtained Z	p value	critical Z
1	128.2	120	40	150	.01	2.51	.0060	2.33
2								

test. As a result, our strategy will be to calculate (1) obtained Z , (2) the associated p -value and (3) critical Z .

Before beginning the calculations, open a new worksheet and, in the first row, enter the numbers 128.2, 120, 40, 150 and .01. These are the function values we will need for our computations. Name the first variable **xbar**, the second **mu**, the third **sigma**, the fourth **n** and the last **alpha**. Your entries should appear as the first five columns in Figure 4.3.

To calculate obtained Z , click **Transform⇒Compute Variable...** The **Compute Variable** window opens. Type *obtained_Z* in the **Target Variable:** window. Enter $(xbar - mu) / (sigma / \sqrt{n})$ in the **Numeric Expression:** window and click **OK**. The variable *obtained_Z* with a value of 2.51 appears in the worksheet.

To calculate the p -value, click **Transform⇒Compute Variable...** Type *p_value* in the **Target Variable:** window. Clear the **Numeric Expression:** window then enter $1 - CDF.NORMAL(obtained_Z, 0, 1)$ in the **Numeric Expression:** window and click **OK**. The variable *p_value* with a value of .01 appears in the Data Editor. Change the number of decimal places for this variable to four. The p -value is now .0060 which is the result reported in the text.

To find critical Z , click **Transform⇒Compute Variable...** Type *critical_Z* in the **Target Variable:** window. Clear the **Numeric Expression:** window then select **Inverse Df** in the **Function group:** window and double-click **Idf.Normal** in the **Functions and Special Variables:** window. Edit this function to read $IDF.NORMAL(1 - alpha, 0, 1)$. This function returns a value such that a specified proportion of the normal curve is below it. The first question mark is replaced by the specified proportion, the second with the mean of the distribution and the third with the standard deviation of the distribution. We specified the mean and standard deviation as zero and one because we are obtaining a Z value from a standard normal distribution. Click **OK**. The value 2.33 is placed in the Data Editor. This is the value reported in the text.

We can now conduct the test of significance by noting that the p -value of .0060 is less than $\alpha = .01$ or noting that obtained Z of 2.51 is greater than critical Z of 2.33.

Task 4.4

Use SPSS to perform the one mean Z test outlined in Example 4.12 on page 97.

Solution

We first note that this task differs from that described in Task 4.3 in that this test is two-tailed while the previous test was one-tailed. We nevertheless follow

the same strategy in that we will calculate (1) obtained Z , (2) the associated p -value and (3) critical Z .

Before beginning the calculations, open a new worksheet and, in the first row, enter the numbers 87.1, 80, 21, 40 and .10. These are the function values we will need for our computations. Name the first variable **xbar**, the second **mu**, the third **sigma**, the fourth **n** and the last **alpha**.

To calculate obtained Z , click **Transform⇒Compute Variable...** The **Compute Variable** window opens. Type *obtained_Z* in the **Target Variable:** window. Enter $(xbar - mu) / (sigma / SQRT(n))$ in the **Numeric Expression:** window and click **OK**. The variable *obtained_Z* with a value of 2.14 appears in the worksheet.

To calculate the p -value, click **Transform⇒Compute Variable...** Type *p_value* in the **Target Variable:** window. Clear the **Numeric Expression:** window then enter $2 * (1 - CDF.NORMAL(obtained_Z, 0, 1))$ in the **Numeric Expression:** window and click **OK**. The variable *p_value* with a value of .0325 appears in the Data Editor which, except for rounding, is the result reported in the text.

To find critical Z , click **Transform⇒Compute Variable...** Type *critical_Z* in the **Target Variable:** window. Clear the **Numeric Expression:** window then select **Inverse Df** in the **Function group:** window and double-click **Idf.Normal** in the **Functions and Special Variables:** window. Edit this function to read $IDF.NORMAL(1 - alpha / 2, 0, 1)$. This function returns a value such that a specified proportion of the normal curve is below it. The first question mark is replaced by the specified proportion, the second with the mean of the distribution and the third with the standard deviation of the distribution. We specified the mean and standard deviation as zero and one because we are obtaining a Z value from a standard normal distribution. Click **OK**. The value 1.64 is placed in the Data Editor. If we increase the number of decimals we obtain a value of 1.645 which, when rounded to 1.65, is the value found in the text. Because the test is two-tailed we know the critical values to be ± 1.65 .

The test of hypothesis can be carried out by noting that the p -value of .0325 is less than $\alpha = .10$ or by noting that obtained Z of 2.14 is greater than the critical Z value of 1.65.

Task 4.5

Use SPSS to perform the one mean t test outlined in Example 4.14 on page 104.

Solution

Enter the data from Example 4.14 into the first column of a new worksheet. Name this variable x . Click **Analyze⇒Compare Means⇒One-Sample T Test...** The **One Sample T Test** window opens. Move x into the **Test Variable(s):** window. Type 8 in the **Test Value:** window. This is the value specified by the null hypothesis. Now click **OK**.

The output gives obtained t as -3.128 which, except for rounding, is the value calculated in the text. The associated p -value is .02. However, this is a

two-tailed value so that the one-tailed value would be $.02/2 = .01$.

Normally, we would not bother performing the critical t versus obtained t version of the test but in the spirit of learning more about SPSS we will next find critical t . To do so, click **Transform**⇒**Compute Variable...** Type *critical_t* in the **Target Variable:** window. Select *Inverse DF* in the **Function group:** window and double-click *Idf. T* in the **Functions and Special Variables:** window. Edit the resultant function to read *IDF. T(.01,6)*. This function returns the t value that has a specified proportion of the t curve below it. The first question mark is replaced by the specified proportion of the curve while the second is replaced by the associated degrees of freedom. Now click **OK**. The value placed in the Data Editor is -3.14 . Because obtained t of -3.128 is greater than this value, the null hypothesis is not rejected.

Notice that this is a rare instance in which the decision *appears* to depend on whether you perform the p -value versus alpha version of the test or the obtained t versus critical t version of the test. Because the reported p -value is $.01$, you would reject on the basis of this criterion. However, SPSS rounded this p -value to $.01$. The actual value is closer to 0.0102012 which would result in failure to reject the null hypothesis. (Are we splitting hairs here?)

Task 4.6

Use SPSS to perform the one sample test for a proportion outlined under the heading **one-tailed test** on page 108.

Solution

SPSS provides a test of the sort needed here but the procedure assumes that calculations are to be made from raw data rather than from data summaries as is given in the text. Our strategy then, is to create a data set with the characteristics described in the problem. To do this, open a new worksheet. Type six 1's and four 0's in the first column. This represents a sample of 10 observations with six (or $.60$) successes. Name this variable x .

Now click **Analyze**⇒**Nonparametric Tests**⇒**Binomial...** The **Binomial Test** window opens. Move the x variable into the **Test Variable List:** window. Be sure the **Get from data** button is selected in the **Define Dichotomy** panel. Type $.38$ in the **Test Proportion** window. Now click the **Exact...** button. Select the **Exact** option and click **Continue** then **OK**.

The reported p -value of $.135$ agrees, given rounding, with the value of $.13475$ calculated in the text. Because this value exceeds $\alpha = .05$, the null hypothesis is not rejected.

Task 4.7

Use SPSS to perform the one sample test for a proportion outlined in Example 4.19 on page 113.

Solution

We will use the same strategy as was employed with Task 4.6. That is, we will create a dataset with the specified characteristics to be submitted for analysis. To this end, open a new worksheet. Type two 1's and six 0's in the first column. This represents a sample of eight observations with two (or .25) successes. Name this variable x .

Now click **Analyze**⇒**Nonparametric Tests**⇒**Binomial...** The **Binomial Test** window opens. Move the x variable into the **Test Variable List:** window. Be sure the **Get from data** button is selected in the **Define Dichotomy** panel. Type .35 in the **Test Proportion** window. Now click the **Exact...** button. Select the **Exact** option and click **Continue** then **OK**.

The reported p -value is .428. But note that this is for a one-tailed test. The two-tailed p -value would be $2 \times .428 = .856$ which agrees, given rounding, with the value of .85562 calculated in the text. Because this value exceeds $\alpha = .05$, the null hypothesis is not rejected.

Task 4.8

Use SPSS to carry out the two-tailed equivalence test outlined in Example 4.25 on page 120.

Solution

Begin by entering the sample data into the first column of a new worksheet. Name this variable x . We can now use SPSS to conduct two one-tailed t tests as is done in the text. However, there is a simpler way to establish two-tailed equivalence which we will also demonstrate.¹ We begin by conducting the tests as was done in the text.

To conduct Test One, click **Analyze**⇒**Compare Means**⇒**One-Sample T Test...** The **One Sample T Test** window opens. Move x into the **Test Variable(s):** window. Type 2 in the **Test Value:** window. This is the value specified by the null hypothesis of Test One. Now click **OK**.

The output gives obtained t as -3.638 which is the value calculated in the text. The associated p -value is .005. However, this is a two-tailed value so that the one-tailed value would be $.005/2 = .0025$ which results in rejection of the null hypothesis for Test One.

Test Two is carried out in the same manner as Test One with -2 replacing 2 in the **Test Value:** window. Obtained t is reported as 2.183 which agrees with that given in the text. The two-tailed p -value is given as 0.057 which would be 0.0285 for the one-tailed test. This results in rejection of the null hypothesis. Because both tests are significant, the null hypothesis of non-equivalence is rejected in favor of the alternative of equivalence.

¹The authors of your text presented two-tailed equivalence testing in the manner they did because they believed that, while more tedious, this method led to a better understanding of the equivalence concept.

A much simpler way of conducting a two-tailed equivalence test² is by forming a $100(1 - 2\alpha)$ percent confidence interval and noting the relationship of L and U to EI_L and EI_U . If both L and U are between EI_L and EI_U , the two-tailed equivalence null hypothesis is rejected otherwise, it is not rejected. You should be able to determine why this is true after you study Section 4.4 and especially Section 4.5. Let us demonstrate this technique on the current example.

To perform the shortcut two-tailed equivalence test, click **Analyze**⇒**Compare Means**⇒**One-Sample T Test...** The **One Sample T Test** window opens. Move x into the **Test Variable(s):** window if it isn't already there. Type 0 in the **Test Value:** window then click the **Options...** button. Type 90 in the **Confidence Interval:** window. This figure is arrived at by noting that $100(1 - 2\alpha) = 100(1 - (2)(.05)) = 90$. Click **Continue** then **OK**. The output gives L as -1.76 and U as $.76$.

Because L of -1.76 is greater than or equal to EI_L of -2 and U of 0.76 is less than or equal to EI_U of 2 , the equivalence null hypothesis is rejected. Had either of these conditions not been met, the equivalence null hypothesis would not have been rejected.

Task 4.9

Use SPSS to carry out the exact two-tailed equivalence test outlined in Example 4.26 on page 122.

Solution

We will use SPSS to test the equivalence null hypothesis by conducting Test One and Test Two as is done in the text. To begin, enter four 1's and four 0's in the first column of a new worksheet. Name this variable x .

To conduct Test One select **Analyze**⇒**Nonparametric Tests**⇒**Binomial...** The **Binomial Test** window opens. Move the x variable into the **Test Variable List:** window. Be sure the **Get from data** button is selected in the **Define Dichotomy** panel. Type $.7$ in the **Test Proportion** window. Now click the **Exact...** button. Select the **Exact** option and click **Continue** then **OK**. The exact one-tailed p -value is reported as $.194$ which, given rounding, agrees with the value calculated in the text.

Conduct Test Two in the same manner as was done for Test One with the exceptions that $.30$ is substituted for $.70$ in the **Test Proportion:** window. The reported p -value is $.194$. Because both of the p -values are less than $\alpha = .20$, both tests are significant which leads to rejection of the null hypothesis of non-equivalence.

²You may wish to delay learning this form of testing until you've studied Sections 4.4 and 4.5.

4.4 Confidence Intervals (page 137)

Task 4.10

Use SPSS to construct the confidence interval described in Example 4.41 on page 147.

Solution

Enter the data from Example 4.41 into the first column of a new worksheet. Name this variable x . Click **Analyze**⇒**Compare Means**⇒**One-Sample T Test....** The **One Sample T Test** window opens. Move x into the **Test Variable(s):** window. Type θ in the **Test Value:** window. Now click **Options...** and type 99 in the **Confidence Interval:** window. Click **Continue** then **OK**. The output gives $L = 94.65$ and $U = 120.15$ both of which agree with results given in the text.

Task 4.11

Use SPSS to form the exact confidence interval constructed in Example 4.42 on page 150.

Solution

While SPSS provides an exact hypothesis test for this problem, it does not provide an exact confidence interval.

Exercises

- 4.1 Use SPSS to reproduce the binomial probabilities in Table 4.5 on text page 109 on page 143.
- 4.2 In Example 4.8 on text page 86, the solution to the problem involves finding the area of the normal curve that lies above a Z value of 1.36. The solution gives this area as .0895. Use SPSS to find this area.
- 4.3 Why was the functional value subtracted from one when finding the p -value in Task 4.3? Why were zero and one used as function input in this problem?
- 4.4 When referring to Task 4.3, how would the function used to find critical Z be changed if the alternative hypothesis had been of the form $H_A : < 120$?
- 4.5 Explain why the expression for finding the p -value in Task 4.4 differs from the expression used in Task 4.3.
- 4.6 Use SPSS to perform the one mean Z test outlined in Example 4.11 on text page 95 as well as to find critical Z for the test.
- 4.7 Use SPSS to perform the one mean Z test outlined in Example 4.13 on text page 98 as well as to find critical Z values for the test.
- 4.8 Use SPSS to perform the one mean t test outlined in Example 4.16 on text page 106.
- 4.9 Use SPSS to perform the test outlined in Example 4.18 on text page 111.
- 4.10 Use SPSS to perform the test outlined in Example 4.20 on text page 114.
- 4.11 Referring to Task 4.8, what would the consequence for Test One have been if obtained t had been 3.638 rather than the value actually obtained of -3.638 . What would the effect be on the test for equivalence?
- 4.12 Use SPSS to perform the following test of hypothesis. HINT: Remember the **Weight Cases...** command?

$$H_{0E} : \pi \leq .45 \text{ or } \pi \geq .55$$

$$H_{AE} : .45 < \pi < .55$$

where $\hat{p} = .47$ and $n = 500$. Use $\alpha = .05$.

- 4.13 Use SPSS to construct the confidence interval outlined in Exercise 4.22 on text page 157.

Chapter 5

Paired Samples Methods

5.1 Introduction

In this chapter you will use SPSS to perform tests of hypotheses and form confidence intervals for various paired samples methods. You will conduct paired samples t tests, McNemar's test, as well as tests on risk and odds ratios. You will also form confidence intervals for some of these statistics and perform equivalence tests.

5.2 Methods Related to Mean Difference (page 160)

Task 5.1

Use SPSS to perform the paired samples t test described in Example 5.1 on page 162. **Solution**

Open the **table 5.1 data.sav** file into a new worksheet. Click **Analyze⇒Compare Means⇒Paired-Samples T Test...** The **Paired-Samples T Test** window opens. Select the **posttreatment** then **pretreatment** variables. When they are both highlighted move them into the **Paired Variables:** window. Click **OK**. The output gives obtained t as 2.931 which agrees, given rounding, with the value calculated in the text. The p -value is .011 which is less than $\alpha = .05$ so that the null hypothesis is rejected.

Task 5.2

Use SPSS to perform the equivalence test described in Example 5.4 on page 168.

Solution

Load the **table 5.5 data.sav** data into a new worksheet. Because the **Paired Samples T Test...** procedure does not permit the user to specify the value of the null hypothesis (it assumes $H_0 : \mu_d = 0$), we will accomplish this task by performing a one mean t test on the difference scores.

To calculate difference scores, click **Transform⇒Compute Variable...** The **Compute Variable** window opens. Type d in the **Target Variable:** window. Move the **lower_dose** variable into the **Numeric Expression:** window, click the minus sign on the numeric keypad, then move the **higher_dose** variable into the **Numeric Expression:** window. Your numeric expression should read $lower_dose - higher_dose$. Click **OK**. The difference scores appear in the Data Editor.

To perform the test of significance, click **Analyze⇒Compare Means⇒One Sample T Test....** The **One Sample T Test** window opens. Move the d variable into the **Test Variable(s):** window. Type 6 in the **Test Value:** window and click **OK**. The output gives obtained t as -2.610 which is the value calculated in the text. The provided p -value is for a two-tailed test so the value for our one-tailed test is $.018/2 = .009$ which is less than $\alpha = .05$ which results in rejection of the null hypothesis.

Task 5.3

Use the paired data in Table 5.1 to construct a one-sided 95 percent confidence interval for the lower bound of μ_d .

Solution

Load the worksheet containing the Table 5.1 data by opening file **table 5.1 data.sav**. Unfortunately, SPSS does not provide a means for constructing one-sided confidence intervals. However, we can take advantage of the fact that the lower end of a two-sided $100(1 - 2\alpha)$ percent confidence interval will be identical to the lower end of a $100(1 - \alpha)$ percent one-sided interval. This means that we can find the lower end of a two-sided $100(1 - (2)(.05)) = 90$ percent interval to solve the problem.¹

To do so, click **Analyze⇒Compare Means⇒Paired-Samples T Test....** The **Paired-Samples T Test** window opens. Select the **posttreat** then **pre-treat** variables. When they are both highlighted move them into the **Paired Variables:** window. Click **Options....** Type 90 in the **Confidence Interval:** window then click **Continue** then **OK**. The output gives the lower limit as 2.553 which, given rounding, is the result reported for Example 5.6

¹If the logic here is not clear, you may wish to review Section 4.5.2 on text page 154.

5.3 Methods Related to Proportions (page 174)

Task 5.4

Perform an exact two-tailed McNemar's test on the data in Table 5.7 on page 175 at the $\alpha = .05$ level.

Solution

To begin, load the worksheet containing the Table 5.7 data which is found in the file **table 5.7 data.sav** into the Data Editor. Notice that the values of the two variables `vaccine_one` and `vaccine_two` are non-numeric. SPSS calls variables of this sort "string" variables. Unfortunately, the procedure that performs McNemar's test uses only numeric data. We will, therefore, have to create new variables of this type.

To do so, click **Transform⇒Recode into Different Variables....** The **Recode into Different Variables** panel opens. Move the `vaccine_one` variable into the **Input Variable—>Output Variable** window. Type `vaccine_one_num` in the **Name:** window of the **Output Variable** panel. Now click **Old and New Values....** The **Recode into Different Variables: Old and New Values** window opens. Be sure the **Value:** button in the **Old Value** panel is selected then type `D_bar` in the accompanying window. Type `1` in the **Value:** window of the **New Value** panel and click **Add**. Repeat the process using `D` in the **Old Value** panel and `0` in the **New Value** panel. The entries '`D_bar`' --> `1` and '`D`' --> `0` are now in the **Old-->New:** window. Now click **Continue** then **Change** then **OK**. Notice that the numeric variable `vaccine_one_num` is now in the Data Editor with 1s and 0s instead of `D_bar` and `D`. We must now create a numeric form for `vaccine_two`.


Click **Transform⇒Recode into Different Variables....** Select `vaccine_one` --> `vaccine_one_num` in the **String Variable -> Output Variable:** window and move it out by clicking the  button. Now move `vaccine_two` into the **Input Variable -> Output Variable:** window then type `vaccine_two_num` in the **Output Variable** window. Click **Old and New Values....** Check to be sure the transformation rules are still in the **Old --> New:** window from the previous transformation. If they are not there, enter them as previously. Click **Continue** then **Change** then **OK**. The variable `vaccine_two_num` is now in the Data Editor. Our Data Editor screen (after a little cleaning up) appears as shown in Figure 5.1. Save this table, we will use it again for Task 5.7. We are now ready to perform the analysis. To conduct McNemar's test, click **Analyze⇒Nonparametric Tests⇒2 Related Samples....** The **Two-Related-Samples Tests** screen opens. Click `vaccine_one_num` then `vaccine_two_num`. When the two variables are highlighted move the pair into the **Test Pair(s) List:**. Check the **McNemar** box and uncheck either of the other two that might be checked. Click **Exact....** The **Exact Tests** window opens. Put a check in the **Exact** box, click **Continue** then **OK**. The output reports the two-tailed p -value as .289 which is the value calculated in conjunction with

Figure 5.1: Data Editor screen after vaccine_one and vaccine_two transformations.

	vaccine_one	vaccine_two	vaccine_one_num	vaccine_two_num
1	D_bar	D	1	0
2	D	D	0	0
3	D_bar	D_bar	1	1
4	D	D_bar	0	1
5	D	D_bar	0	1
6	D_bar	D_bar	1	1
7	D	D_bar	0	1
8	D	D	0	0
9	D_bar	D_bar	1	1
10	D_bar	D	1	0
11	D	D_bar	0	1
12	D_bar	D_bar	1	1
13	D	D_bar	0	1
14	D_bar	D_bar	1	1
15	D_bar	D_bar	1	1
16	D	D_bar	0	1

Example 5.11 on page 179 of the text

Task 5.5

Use the Outcome variable in Table 5.14 to perform the exact test outlined in Example 5.13 on page 183.

Solution

Enter the Outcome variable, minus the noninformative observations, from Table 5.14 into the first column of a new worksheet. Name this variable **outcome**. Click **Analyze**⇒**Nonparametric Tests**⇒**Binomial...** The **Binomial Test** window opens. Move the outcome variable into the **Test Variable List:** window. Type **.4** in the **Test Proportion:** window then click **Exact...** Choose **Exact** in the **Exact Tests** panel then click **Continue** then **OK**. The output gives the one-tailed p -value² as .514 which is the value calculated in the solution to Example 5.13.

Task 5.6

Construct the confidence intervals described in Example 5.15 on page 186.

Solution

Unfortunately, SPSS does not provide a simple means of constructing such intervals.

²This is for an upper-tailed test. SPSS informs you when reporting the p -value for a lower-tailed test.

5.4 Methods Related to Paired Samples Risk Ratios (page 190)

Task 5.7

Carry out the tasks outlined in Example 5.17 on page 191.

Solution

SPSS does not provide a direct means for calculating a paired samples risk ratio. However, it can still be useful in such an endeavor in that it can be used to form the data into a two by two table thereby facilitating the calculation. This can be particularly useful for large data sets.

For the problem at hand, load file **table 5.7 data.sav** into the data editor. Remember that you created the variables `vaccine_one_num` and `vaccine_two_num` in conjunction with Task 5.4 and stored them in this file. Now click **Analyze**⇒**Tables**⇒**Basic Tables...** The **Basic Tables** panel opens. Move the `vaccine_two_num` variable into the **Down:** window and the `vaccine_one_num` variable into the **Across:** window. Click **OK**. The output provides a two by two table with cell entries as shown in the solution to Example 5.17 on page 191. You can now use a hand held calculator to calculate the risk ratio via Equation 5.7 on page 191. The test of significance conducted via Equation 5.4 was carried out in the solution to Task 5.4. To obtain the result provided by Equation 5.5, simply square the result given by Equation 5.4.

Task 5.8

Use SPSS to perform the equivalence test as outlined in Example 5.20 on page 195.

Solution

Unfortunately, SPSS does not provide a straightforward means for performing this test.

Task 5.9

Use SPSS to form the confidence interval described in Example 5.22 on page 197.

Solution

Unfortunately, SPSS does not provide a straightforward means for constructing this interval.

Figure 5.2: Data Editor screen showing weighted data used for McNemar test.

	cancer	no_cancer	frequency
1	1	1	13
2	1	0	25
3	0	1	5
4	0	0	55
5			

5.5 Methods Related to Paired Samples Odds Ratios (page 199)

Task 5.10

Use SPSS to perform exact and approximate tests of the hypothesis $H_0 : OR = 1$ as described in Example 5.24 on page 201. ³ **Solution**

Using the fact that a test of $H_0 : OR = 1$ is equivalent to a test of $H_0 : \pi = .5$ which is the hypothesis addressed by the McNemar test, we will perform the analysis by means of the McNemar procedure. In order to do so, we will first enter the data for the analysis.

To do so, open a new worksheet and enter the data as shown in Figure 5.2. Notice that the data pairs, 1 1, 1 0, 0 1 and 0 0 are used to represent the outcomes; both pair members exposed, cancer worker exposed with non-cancer worker un-exposed, cancer worker un-exposed and non-cancer worker exposed, and both workers un-exposed. The third column gives the number of pairs in each category and will be used to weight the cases as was done for Task 2.10.

After entering the data and naming the variables, click **Data⇒Weight Cases....** The **Weight Cases** window opens. Select the **Weight cases by** button and move the frequency variable into the **Frequency Variable:** window and click **OK**. From the SPSS perspective, there are now 13 1 1 data pairs, 25 1 0 data pairs etc. in the Data Editor. We can now perform the analysis.

To perform the McNemar test, click **Analyze⇒Nonparametric Tests⇒2 Related Samples....** The **Two-Related-Samples Tests** window opens. Click on the **cancer** then **no_cancer** variables and move the variable pair into the **Test Pair(s) List:** window. Check the **McNemar** box and uncheck either of the other two boxes that might be checked. Click **Exact...** then choose the **Exact** option in the **Exact Tests** panel. Click **Continue** then **OK**.

The output gives the exact two-tailed p -value as zero to three decimal places. The chi-square value is given as 12.033 which differs from the 13.33 calculated in the text. The difference is attributable to a continuity correction applied

³There is an error in the text on this page. See the errata sheet for the text at <http://www.biostats-hs.com>.

to the calculation by SPSS that is not utilized in the text. Corrections of this sort result in a smaller chi-square value than does the uncorrected chi-square statistic.

Task 5.11

Perform the exact two-tailed equivalence test described in Example 5.27 on page 206.

Solution

Unfortunately, SPSS does not provide a straightforward means of performing this test.

Task 5.12

Construct the exact confidence interval described in Example 5.30 on page 211.

Solution

Unfortunately, SPSS does not provide a straightforward means of constructing this interval.

Exercises

- 5.1 Use SPSS to perform the test of hypothesis outlined in Example 5.2 on page 163.
- 5.2 Use SPSS to perform the equivalence test outlined in Example 5.5 on page 169.
- 5.3 Referring to Task 5.2, what would you have done differently had the required test been Test Two rather than Test One?
- 5.4 Use the paired data in Table 5.2 on page 163 to form a 95 percent confidence interval for the estimation of $\mu_1 - \mu_2$ where μ_1 represents the mean for Treatment One and μ_2 represents the mean for Treatment Two.
- 5.5 Use SPSS to perform the test outlined in Example 5.12 on page 180.
- 5.6 Use the Outcome variable in Table 5.16 to perform the exact test alluded to in Example 5.14 on page 185.
- 5.7 Conduct the test outlined in Example 5.18 on page 192.
- 5.8 Conduct the test of significance alluded to in Example 5.19 on page 192.
- 5.9 Use SPSS to perform the exact two-tailed test of significance outlined in Example 5.25 on page 202.

Chapter 6

Two Independent Samples Methods

6.1 Introduction

In this chapter you will use SPSS to perform tests of hypotheses and form confidence intervals for various two (independent) sample methods. You will conduct independent samples t tests and tests for differences between proportions. You will also form confidence intervals for each of these statistics and perform equivalence tests.

6.2 Methods Related to Differences Between Means (page 215)

Task 6.1

Conduct the independent samples t test described in Example 6.1 on page 220.

Solution

Load the **table 6.1 data.sav** worksheet into the Data Editor. Notice that the outcome variable is named Outcome and is in the second column. The first column named Group contains a designator as to which group each outcome value belongs. Thus, the first value in column two is 129 and is for a subject in group one while the 16th value is 138 and is associated with group two..

To conduct the t test, click **Analyze⇒Compare Means⇒Independent-Samples T Test....** The **Independent-Samples T Test** window opens. Move the **Outcome** variable into the **Test Variable(s):** window then move the **Group** variable into the **Grouping Variable:** window. Now click the **Define Groups...** button. Be sure the **Use specified values** button is selected then

Figure 6.1: SPSS output for an independent samples t test.

Group Statistics									
Group		N	Mean	Std. Deviation	Std. Error Mean				
Outcome	1	15	131.80	12.531	3.236				
	2	15	135.20	10.421	2.691				

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	L
Outcome	Equal variances assumed	.274	.605	-.808	28	.426	-3.400	4.208	-.1
	Equal variances not assumed			-.808	27.099	.426	-3.400	4.208	-.1

type 1 in the **Group 1:** window and 2 in the **Group 2:** window. Click **Continue** then **OK**. The output is shown in Figure 6.1.

The portion of the output labeled **Group Statistics** shows the number of observations in each group (N), along with the mean, standard deviation and standard error of the mean for each group. The portion labeled **Independent Samples Test** gives the results of two types of t tests as well as a test for equality of variances for the two groups. The F value of .274 with the associated p -value of .605 are the results of a test of the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ where σ_1^2 and σ_2^2 are the variances of the populations from which the two samples were drawn. You will recall that one of the assumptions underlying the independent samples t test is equality of population variances. This, then, is a test of that assumption.¹

Labeling for the two types of t tests are “Equal variances assumed” which is the test discussed in the text and “Equal variances not assumed” which was not discussed in the text. As may be seen, obtained t is $-.808$ which is the value calculated in the text. The associated two-tailed p -value of .426 results in failure to reject the null hypothesis of identical population means. Notice that a confidence interval for the estimation of $\mu_1 - \mu_2$ is also provided.

Task 6.2

Conduct the equivalence test described in Example 6.3 on page 225.

Solution

Unfortunately, SPSS does not provide a direct means to test $H_0 : \mu_1 - \mu_2 = -5$ which is required for Test Two. Rather, it insists on testing $H_0 : \mu_1 - \mu_2 = 0$. We can circumvent this difficulty in several ways. We will demonstrate one of these here and leave a second as an exercise.

¹Can it be asserted that the population variances are equal if this test is non-significant?

Suppose that $\mu_1 - \mu_2$ is exactly equal to -5 . If we add five points to each score in group one, we can now conceive of this group as having been sampled from a population whose mean is five points higher than previously so that $\mu_1 - \mu_2$ is now zero. Rejecting the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ after adding the points is the same thing as rejecting $H_0 : \mu_1 - \mu_2 = -5$ for the original data. So this will be our strategy.

Load the **table 6.1 data.sav** worksheet into the Data Editor if it isn't already there. To add the five points and create a new variable click **Transform**⇒**Compute Variable....** The **Compute Variable** window opens. Type *Outcome5* in the **Target Variable:** window then enter *Outcome + 5* in the **Numeric Expression:** window. This series of commands will add five points to all Outcome scores. But we want to add five points only to Outcome scores associated with group one. To restrict our transformation to group one, click the **if...** button. The **Compute Variable: If Cases** window opens. Select the **Include if case satisfies condition:** option. Now enter *Group = 1* in the associated window. To this point we have instructed SPSS to create a new variable named Outcome5 which is defined as Outcome plus five points for all cases that belong to group one. Click **Continue** then **OK**. The variable Outcome5 appears in the third column of the Data Editor. While the values of Outcome5 are five points greater than those for variable Outcome for subjects in the first group, the value for group two are missing. That is because we gave SPSS no instructions regarding the scores of group two subjects.

To create Outcome5 scores for group two click **Transform**⇒**Compute Variable....** The **Compute Variable** window opens. Type *Outcome5* in the **Target Variable:** window then enter *Outcome* in the **Numeric Expression:** window. Click the **if...** button. The **Compute Variable: If Cases** window opens. Select the **Include if case satisfies condition:** option. Now enter *Group = 2* in the associated window. Click **Continue** then **OK**. SPSS will query, "Change existing variable?" Click **OK**. The missing values of Outcome5 for group two are now filled in with scores from the variable Outcome. We are now ready to perform the analysis.

To conduct the analysis, click **Analyze**⇒**Compare Means**⇒**Independent-Samples T Test....** The **Independent-Samples T Test** window opens. Move the Outcome5 variable into the **Test Variable(s):** window. Then move the Group variable into the **Grouping Variable:** window. Click the **Define Groups:** button. Select the **Use specified values** option then type *1* in the **Group 1:** window and *2* in the **Group 2:** window. Click **Continue** then **OK**. The output gives obtained *t* as .380 which is the same value calculated in the text. The *p*-value of .707 is for a two-tailed test. In order to obtain the *p*-value for our one-tailed test we calculate $.707/2 = .354$ so that the equivalence null hypothesis is not rejected.

Task 6.3

Form the confidence interval described in Example 6.5 on page 228.

Figure 6.2: Data for Task 6.4.

	group	outcome	frequency
1	1	1	23
2	1	2	291
3	2	1	39
4	2	2	277

Solution

Load the **table 6.1 data.sav** worksheet into the Data Editor if it isn't already there. To form the interval, click **Analyze**⇒**Compare Means**⇒**Independent-Samples T Test....** The **Independent-Samples T Test** window opens. Move the **Outcome** variable into the **Test Variable(s):** window then move the **Group** variable into the **Grouping Variable:** window. Click the **Define Groups:** button. Select the **Use specified values** option then type *1* in the **Group 1:** window and *2* in the **Group 2:** window. Click **Continue**. Now click the **Options...** button. The **Independent-Samples T Test: Options** window opens. Type *95* in the **Confidence Interval:** window if it isn't already there. Click **Continue** then **OK**. The output gives *L* and *U* as -12.020 and 5.220 respectively. Given rounding, both figures agree with the computations in the text.

6.3 Methods Related to Proportions (page 230)

Task 6.4

Perform the two sample test for a difference between proportions addressed in Example 6.7 on page 233.

Solution

Enter the data shown in Figure 6.2 of this manual into a new worksheet. The entries under the **group** variable are 1 for membership in the instruction group and 2 for membership in the no instruction group. Entries for the **outcome** variable are 1 for a low birth weight baby and 2 for a baby that is not low birth weight. The **frequency** column gives the frequency with which each group/outcome combination occurred in the study.

To begin, click **Data**⇒**Weight Cases....** The **Weight Cases** window opens. Move the **frequency** variable into the **Frequency Variable:** window. Click **OK**. From the SPSS point of view, there are now 23 group 1, outcome 1 pairs in the data set, 291 group 1 outcome 2 pairs and so on.

To perform the analysis click **Analyze**⇒**Descriptive Statistics**⇒**Crosstabs....** The **Crosstabs** window opens. Move the **group** variable into the **Row(s):** window and the **outcome** variable into the **Column(s):** window.

Now click the **Statistics...** button. The **Crosstabs: Statistics** window opens. Put a check in the **Chi-Square** box then click **Continue** then **OK**. The output gives obtained chi-square as 4.468. Given rounding, this is the square of the Z value calculated in the text. The p -value is given as .035 which would be appropriate for a two-tailed test of the hypothesis $H_0 : \pi_1 = \pi_2$. Because the current problem calls for a one-tailed test we calculate $.035/2 = .0175$ as the appropriate p -value.

Task 6.5

Perform the two-tailed equivalence test described in Example 6.9 on page 235.

Solution

Unfortunately, SPSS does not provide a procedure for conducting this test.

Task 6.6

Construct the confidence interval described in Example 6.11 on page 236.

Solution

Unfortunately, SPSS does not provide a procedure for constructing this interval.

6.4 Methods Related to Independent Samples Risk Ratios (page 238)

Unfortunately, SPSS does not implement the methods discussed in the text for dealing with independent samples risk ratios.

6.5 Methods Related to Independent Samples Odds Ratios (page 247)

Unfortunately, SPSS does not implement the methods discussed in the text for dealing with independent samples odds ratios.

Exercises

- 6.1 Use the summary values reported in the SPSS output associated with Task 6.1 (i.e. Figure 6.1 on page 58 of this manual) to calculate the obtained t value of $-.808$ reported in that output.
- 6.2 Use SPSS to conduct the test of hypothesis described in Example 6.2 on text page 221. What is the appropriate p -value for this test?
- 6.3 Use SPSS to perform the equivalence test described in Example 6.4 on page 226 of the text.
- 6.4 Use SPSS to form the confidence interval discussed in Example 6.6 on page 229 of the text.
- 6.5 Use SPSS to perform the test described in Example 6.8 on page 233 of the text.

Chapter 7

Multi-Sample Methods

7.1 Introduction

In this chapter you will use SPSS to perform a One-Way ANOVA, a 2 by k chi-square test and Tukey's HSD test.

7.2 The One-Way Analysis of Variance (ANOVA) F Test (page 264)

Task 7.1

Use the data in Table 7.1 on page 266 of the text to compute MS_w , MS_b , obtained F and to produce an ANOVA table.

Solution

Load the **table 7.1 data.sav** data into the Data Editor. Notice that the **Response** variable contains the weights of the 15 subjects while the **Group** variable consists of an indicator (1, 2, or 3) as to which group the weight represents.

To perform the analysis, click **Analyze⇒Compare Means⇒One-Way ANOVA....** The **One-Way ANOVA** window opens. Move the **Response** variable into the **Dependent List:** window and the **Group** variable into the **Factor:** window.

Though we are not asked to do so in this Task, let's go ahead and perform Tukey's HSD test at this time. To do so, click **Post Hoc....** The **One-Way ANOVA: Post Hoc Multiple Comparisons** window opens. Place a check in the **Tukey** box then click **Continue** and **OK**. The relevant portion of the SPSS output is shown in Figure 7.1. Notice that, except for rounding, these are the results reported in Table 7.4 on page 270 of the text. The p -value is reported as .263.

Figure 7.1: SPSS rendering of ANOVA table for data in text Table 7.1.

ANOVA					
Response					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1869.733	2	934.867	1.496	.263
Within Groups	7500.000	12	625.000		
Total	9369.733	14			

Figure 7.2: Results of Tukey's HSD test conducted on data in text Table 7.1.

Multiple Comparisons						
Dependent Variable: Response						
Tukey HSD						
(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-7.800	15.811	.876	-49.98	34.38
	3	18.800	15.811	.482	-23.38	60.98
2	1	7.800	15.811	.876	-34.38	49.98
	3	26.600	15.811	.251	-15.58	68.78
3	1	-18.800	15.811	.482	-60.98	23.38
	2	-26.600	15.811	.251	-68.78	15.58

A portion of the output relevant to the Tukey test is shown in Figure 7.2. The designations (I) Group and (J) Group show the groups being compared. For example, group 1 is being compared with groups 2 and 3 in the first block, group 2 is being compared with groups 1 and 3 in the second block and so forth. The comparisons are done in the form of confidence intervals rather than hypothesis tests but the reported p -values (as well as the presence of zero in each interval) which are labeled Sig. in the output, indicate that none of the comparisons are significant which is the result obtained in the text calculations on page 290.

7.3 The 2 By k Chi-Square Test (page 276)

Task 7.2

Perform the chi-square test outlined in Example 7.6 on page 280.

Solution

Load the **example 7.6 data.sav** data into the Data Editor. Notice that the variable **group** is 1 for those receiving nutritional instruction and 2 for those not receiving such instruction. Likewise, the variable **weight** is 1 for low birth weight babies and 2 for normal weight babies. The **frequency** variable shows the number of subjects in each group/weight category combination. We must

begin by informing SPSS that the **frequency** variable indicates the number of group/weight combinations in the dataset.

To do so, click **Data⇒Weight Cases...** The **Weight Cases** window opens. Choose the **Weight cases by** option then move the **frequency** variable into the **Frequency Variable:** window. Click **OK**. SPSS now understands that there were 23 low birth weight babies in the group receiving instruction, 291 normal birth weight babies in the group receiving instruction etc.

To perform the analysis click **Analyze⇒Descriptive Statistics⇒Crosstabs...** The **Crosstabs** window opens. Move the **weight** variable into the **Row(s):** window and the **group** variable into the **Column(s):** window. Click the **Statistics...** button. The **Crosstabs: Statistics** window opens. Put a check in the **Chi-Square** box and click **Continue** then **OK**.

SPSS gives the obtained (Pearson) Chi-Square value as 4.468 which agrees with the result obtained in the text. The associated p -value is given as .035 which results in rejection of the null hypothesis at $\alpha = .05$. This is the result obtained in the text using the obtained versus critical chi-square method.

7.4 Multiple Comparison Procedures (page 282)

Task 7.3

Perform the Tukey's HSD tests outlined in Example 7.9 on page 289.

Solution

See Task 7.1 on page 63 of this manual.

Exercises

- 7.1 Carry out the ANOVA analysis outlined in Example 7.3 on page 269 of the text.
- 7.2 Carry out the ANOVA analysis outlined in Example 7.4 on page 271.
- 7.3 Perform the chi-square analysis outlined in Example 7.5 on page 279.
- 7.4 Perform the multiple comparison tests outlined in Example 7.10 on page 290.

Chapter 8

The Assessment of Relationships

8.1 Introduction

In this chapter you will use SPSS to calculate the Pearson product-moment correlation coefficient and to test the hypothesis $H_0 : \rho = 0$. Additionally, you will conduct the chi-square test for independence.

8.2 The Pearson Product-Moment Correlation Coefficient (page 295)

Task 8.1

Calculate the Pearson product-moment correlation coefficient for the data in Table 8.1 on page 297. Perform a two-tailed test of the hypothesis $H_0 : \rho = 0$ at the .05 level.

Solution

Load the **table 8.1 data.sav** worksheet into the Data Editor. Click **Analyze**⇒**Correlate**⇒**Bivariate....** The **Bivariate Correlations** window opens. Move the **x** and **y** variables into the **Variables:** window. Be sure the **Pearson** and **Two-tailed** options are selected then click **OK**. The resultant output gives the correlation coefficient as .964 which is the value calculated in the text. The two-tailed p -value of .000 results in reject of the null hypothesis $H_0 : \rho = 0$.

Figure 8.1: Chi-square analysis of data described in Example 8.7.

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	206.811 ^a	4	.000
Likelihood Ratio	217.545	4	.000
Linear-by-Linear Association	131.554	1	.000
N of Valid Cases	3095		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 114.60.

8.3 The Chi-Square Test For Independence (page 312)

Task 8.2

Perform the chi-square test described in Example 8.7 on page 314.

Solution

Load the **example 8.7 data.sav** file into the Data Editor. Notice that the numbers 1, 2 and 3 of the **county** variable designate the three counties while the numbers 1, 2 and 3 of the **status** variable indicate respectively “vaccinated,” “not vaccinated” and “unknown vaccination status.” As we have done before, we must inform SPSS of the number or frequency of each county/status combination. This is the observed frequency for the chi-square analysis.

To do so, click **data⇒Weight Cases...** The **Weight Cases** window opens. Choose the **Weight cases by** option and move the **frequency** variable into the **Frequency Variable:** window. Click **OK**.

To perform the chi-square analysis click **Analyze⇒Descriptive Statistics⇒Crosstabs...** The **Crosstabs** window opens. Move the **county** and **status** variables into the **Rows(s):** and **Columns(s):** windows respectively. Click the **Statistics...** button and put a check in the **Chi-square** box. Click **Continue**. Now click **Cells...** and put checks in the **Observed** and **Expected** boxes. Click **Continue** then **OK**. The relevant portion of the output is shown in Figure 8.1. As may be seen, the observed and expected frequencies for each cell of the table are printed along with the obtained chi-square value of 206.811. The associated *p*-value is 0.000. This agrees with results given in the text.

Exercises

8.1 Calculate the correlation coefficient discussed in Example 8.2 on page 298.

Chapter 9

Linear Regression

9.1 Introduction

In this chapter, you will use SPSS to construct simple and multiple regression models, generate \hat{R}^2 values and conduct tests of significance.

9.2 Simple Linear Regression (page 320)

Task 9.1

Construct the simple linear regression model alluded to in Example 9.1 on page 321.

Solution

Open the **table 8.2 data** worksheet. Click **Analyze**⇒**Regression**⇒**Linear....** The **Linear Regression** window opens. Move the **y** variable into the **Dependent:** window then move **x** into the **Independent(s):** window. Click **OK**. Under the B column of the printout, SPSS gives the regression equation as $\hat{y} = .874 + .706x$ which is the result given in the text.

Task 9.2

Find SS_{reg} , SS_{res} , and \hat{R}^2 for the data in Table 9.1 on page 323.

Solution

Open the **table 9.1 data.sav** worksheet. Click **Analyze**⇒**Regression**⇒**Linear....** The **Linear Regression** window opens. Move the **y** variable into the **Dependent:** window then move **x** into the **Independent(s):** window. Click **Statistics....** Put check marks in the **Estimates**, **Model fit** and **R squared change** windows. Click **Continue** then **OK**.

The output gives the regression and residual sum of squares as 107.444 and 8.290 respectively which agrees with the values calculated in the text. \widehat{R}^2 is given as .928 which is the value calculated in the solution to Example 9.2 on page 324 of the text.

Task 9.3

Test the null hypothesis $R^2 = 0$ for a model used to predict y from x for the data in Table 9.1 on page 323.

Solution

Using the output from the solution to Task 9.2, we see that obtained F is reported as 168.494 with an associated p -value of .000. The value of 168.494 reported by SPSS differs slightly from the value of 167.56 calculated in the text. This small difference appears to be the result of differences in rounding.

9.3 Multiple Linear Regression (page 329)

Task 9.4

Use the data in Table 9.2 on page 330 to construct a two predictor model with x_1 and x_2 being used to predict y .

Solution

Open the **table 9.2 data.sav** worksheet. Click **Analyze**⇒**Regression**⇒**Linear...** The **Linear Regression** window opens. Move the **y** variable into the **Dependent:** window then move **x1** and **x2** into the **Independent(s):** window. Click **Statistics....** Put check marks in the **Estimates** and **Model fit** windows. Click **Continue** then **OK**. SPSS gives the regression model as $\hat{y} = 2.686 + .404x_1 + 1.00x_2$ which agrees with the results calculated in the text.

Task 9.5

Test the null hypothesis $H_0 : \beta_1 = \beta_2 = 0$ for the model you constructed in Task 9.4.

Solution

From the output for the solution to Task 9.4, we note that obtained F for the regression model is 4.725 which agrees with the value calculated in the text. The associated p -value is 0.031 which results in rejection of the null hypothesis for a test conducted at $\alpha = .05$. This is the same conclusion reached in the text.

Task 9.6

Use the data in Table 9.2 on page 330 to determine whether adding x_2 to a

model that contains x_1 adds significantly to \widehat{R}^2 .

Solution

This problem can be approached in a number of different ways. For example, we can note from the output associated with Task 9.4 that the p -value associated with the t statistic for the test $H_0 : \beta_2 = 0$ is .084. This is the same p -value as would be obtained for a test of $H_0 : \widehat{R}_{12}^2 - \widehat{R}_1^2 = 0$.

For a second way, load the **table 9.2 data.sav** data file into the Data Editor if it isn't already there. Click **Analyze**⇒**Regression**⇒**Linear...** The **Linear Regression** window opens. Move the **y** variable into the **Dependent:** window. Move **x1** into the **Independent(s):** window and click **Next**. This assures that **x1** will enter the equation first. Now move **x2** into the **Independent(s):** window and click **Next** again. To have the R square values printed click **Statistics...** and check the **R squared change** box. Click **Continue** then **OK**.

Notice that the R square change when x_2 entered the model was .166. The F value used to test whether this change differs from zero was 3.551 which has an associated p -value of .084 which is the same value obtained previously for a test of $H_0 : \beta_2 = 0$.

Exercises

- 9.1 Find the regression equation for predicting y from x for the data in Table 9.1 on page 323.
- 9.2 Use SPSS to construct the model described in Exercise 9.7 on page 341 of the text.
- 9.3 Test the null hypothesis $H_0 : R^2 = 0$ for the model you constructed in Exercise 9.2.
- 9.4 Use the data in Table 9.2 on page 330 to determine whether adding x_1 to a model that contains x_2 adds significantly to \hat{R}^2 .

Chapter 10

Methods Based on the Permutation Principle

10.1 Introduction

In this chapter you will use SPSS to perform Wilcoxon's signed-ranks test, Wilcoxon's rank-sum (Mann-Whitney) test, the Kruskal-Wallis test and Fisher's exact test.

10.3 Applications (page 348)

Task 10.1

Perform the Wilcoxon's signed-ranks test described in Example 10.15 on page 369.

Solution

Open the **table 10.7 data.sav** data file. Click **Analyze⇒Nonparametric Tests⇒2 Related Samples....** The **Two-Related-Samples Tests** window opens. Select the **post_treatment** and **pre_treatment** variables and move them into the **Test Pair(s) List:** window. Under **Test Type** check **Wilcoxon**. Now click the **Exact...** button. Choose the **Exact** option, click **Continue** then **OK**. The p -value for **Exact Sig. (2 -tailed)** is given as .275 which results in failure to reject the null hypothesis. This is the conclusion reported in the text.

Task 10.2

Perform the Wilcoxon's rank-sum test described in Example 10.20 on page 382.

Solution

Open the **example 10.20 data.sav** data file. Click **Analyze⇒Nonparametric Tests⇒2 Independent Samples....** The **Two-Independent-Samples Tests** window opens. Place a check in the **Mann-Whitney U** box.¹ Click the **Exact...** button and select the **Exact** option. Click **Continue**. Move the **outcome** variable into the **Test Variable List:** window and the **group** variable into the **Group Variable:** window. Now click the **Define Groups...** button. Type *1* in the **Group 1:** box and *2* in the **Group 2:** box. Click **Continue** then **OK**. The exact one-tailed p -value is given as .350 which is the value calculated on text page 382.

Task 10.3

Perform a Kruskal-Wallis test on the raw data in the table on page 393.

Solution

Enter the group number, i.e. 1, 2, or 3, of each of the 6 subjects in the first column of a new work sheet. Name this variable “group.” Enter the outcome measures in the second column and name this variable “outcome.” Click **Analyze⇒Nonparametric Tests⇒K Independent Samples....** The **Tests for Several Independent Samples** window opens. Move the **outcome** variable into the **Test Variable List:** window and the **group** variable into the **Grouping Variable:** window. Click the **Define Range...** button. Enter *1* in the **Minimum** and *3* in the **Maximum:** window. Click **Continue**. Check the **Kruskal-Wallis H** box then click the **Exact...** button. Choose the **Exact** option then click **Continue** and then **OK**. The exact p -value is given as .20 which is the value calculated in the text.

Task 10.4

Perform a two-tailed Fisher’s exact test on the data in Table 10.23 on page 405.

Solution

Load **table 10.23 data.sav** into the Data Editor. We begin by weighting the cell identifiers by the frequencies. That is, we must tell SPSS how many observations are in each of the cells. To do this, click **Data⇒Weight Cases....** The **Weight Cases** window opens. Choose the **Weight cases by** option and move the **frequency** variable into the **Frequency Variable:** window. Click **OK**. We are now ready to perform Fisher’s exact test.

To do so, Click **Analyze⇒Descriptive Statistics⇒Crosstabs....** The **Crosstabs** window opens. Move the **smoke** variable into the **Row(s):** window and the **disease** variable into the **Column(s):** window. Click **Statistics...**, place a check in the **Chi-square** box then click **Continue**. Now click **Exact...**,

¹Recall that the Mann-Whitney U test produces the same result as does the Wilcoxon rank-sum test.

choose the **Exact** option and click **Continue**. Now click **OK**.

The output gives the exact two-tailed p -value for Fisher's exact test as .089.. This is the value calculated for a two-tailed test in the solution to Example 10.30 on text page 405.

Exercises

- 10.1 Conduct the test described in Example 10.16 on page 371.
- 10.2 Perform the test alluded to in Example 10.21 on page 383.
- 10.3 Conduct the test described in Example 10.22 on page 385.
- 10.4 Conduct the Kruskal-Wallis test described in Example 10.27 on page 395.
- 10.5 Conduct the test described in Example 10.31 on page 407.

Bibliography

- [1] Arthur Griffith. *SPSS for Dummies*. Wiley Publishing Inc., Hoboken, New Jersey, 2007.
- [2] Marija J. Norušis. *SPSS 15.0 Guide to Data Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 2006.