Finally, use the skewness and the standard error of the skewness to find a z-score:

$$z_{S_k} = \frac{S_k - 0}{SE_{S_k}} = \frac{-1.018}{0.501}$$

$$z_{S_k} = -2.032$$

Use the z-score to examine the sample's approximation to a normal distribution. This value must fall between -1.96 and $+1.96$ to pass the normality assumption for $\alpha = 0.05$. Since this z-score value does not fall within that range, the sample has failed our normality assumption for skewness. Therefore, either the sample must be modified and rechecked or you must use a nonparametric statistical test.

2.4.3   Examining Skewness and Kurtosis for Normality Using SPSS

We will analyze the examples earlier using SPSS.

2.4.3.1   Define Your Variables   First, click the "Variable View" tab at the bottom of your screen. Then, type the name of your variable(s) in the "Name" column. As shown in Figure 2.7, we have named our variable "Wk1_Qz."
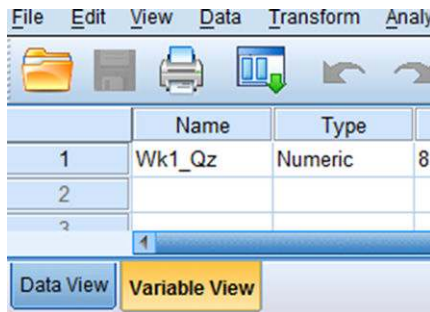

FIGURE 2.7

2.4.3.2   Type in Your Values   Click the "Data View" tab at the bottom of your screen and type your data under the variable names. As shown in Figure 2.8, we have typed the values for the "Wk1_Qz" sample.

2.4.3.3   Analyze Your Data   As shown in Figure 2.9, use the pull-down menus to choose "Analyze," "Descriptive Statistics," and "Descriptives . . ."

Choose the variable(s) that you want to examine. Then, click the button in the middle to move the variable to the "Variable(s)" box, as shown in Figure 2.10. Next, click the "Options . . ." button to open the "Descriptives: Options" window shown in Figure 2.11. In the "Distribution" section, check the boxes next to "Kurtosis" and "Skewness." Then, click "Continue."

Finally, once you have returned to the "Descriptives" window, as shown in Figure 2.12, click "OK" to perform the analysis.
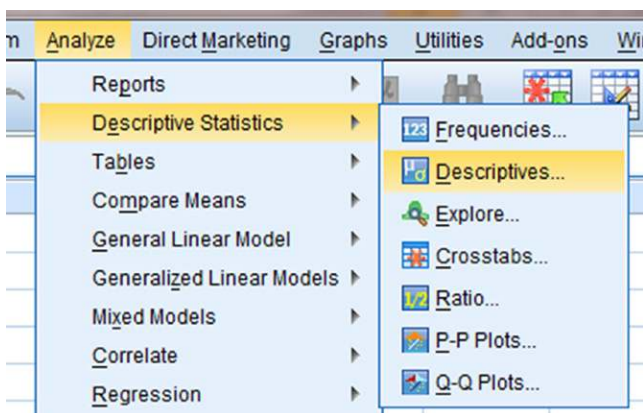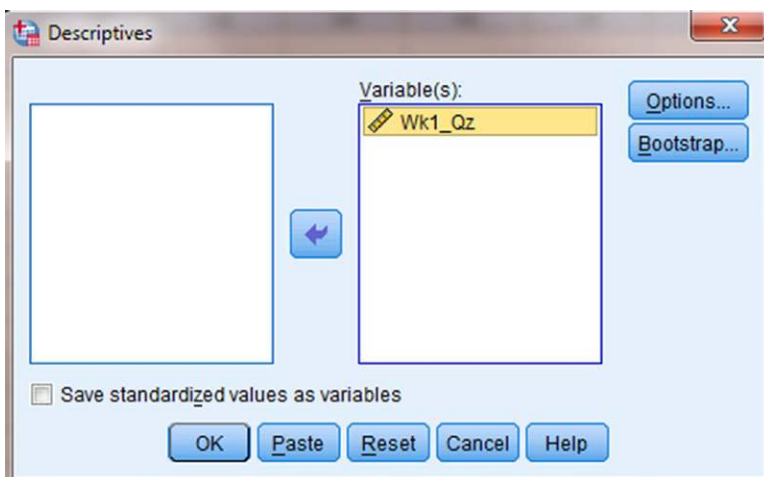
FIGURE 2.8

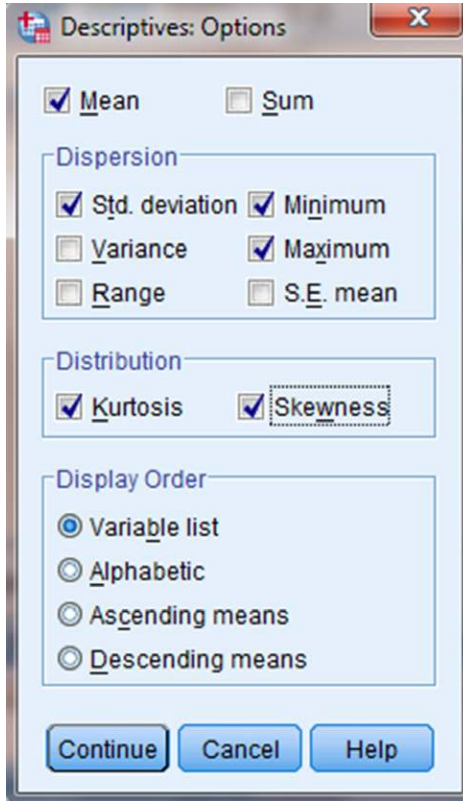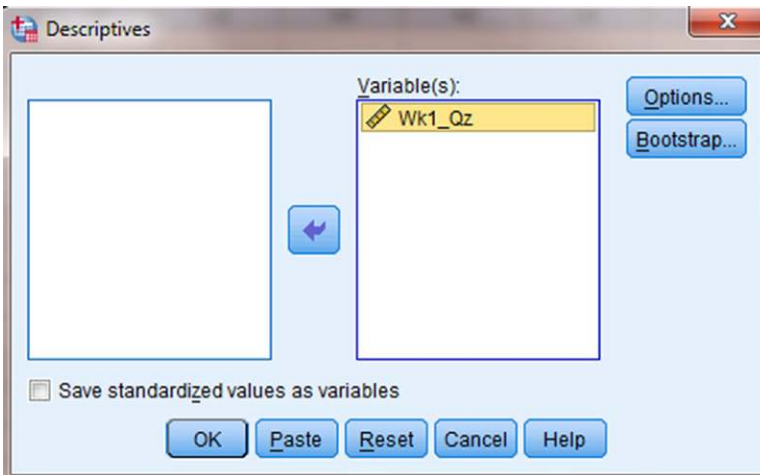**FIGURE 2.9**

FIGURE 2.10

FIGURE 2.11

FIGURE 2.12

2.4.3.4 *Interpret the Results from the SPSS Output Window*   The SPSS Output 2.1 provides the kurtosis and the skewness, along with their associated standard errors. In our example, the skewness is -1.018 and its standard error is 0.501. The kurtosis is 1.153 and its standard error is 0.972.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Wk1_Qz	21	35.00	100.00	80.2381	16.62199	-1.018	.501	1.153	.972
Valid N (listwise)	21								

SPSS OUTPUT 2.1

At this stage, we need to manually compute the z-scores for the skewness and kurtosis as we did in the previous examples. First, compute the z-score for kurtosis:

$$z_K = \frac{K-0}{SE_K} = \frac{1.153-0}{0.972}$$

$$z_K = 1.186$$

Next, we compute the z-score for skewness:

$$z_{S_k} = \frac{S_k - 0}{SE_{S_k}} = \frac{-1.018}{0.501}$$

$$z_{S_k} = -2.032$$

Both of these values must fall between -1.96 and $+1.96$ to pass the normality assumption for $\alpha = 0.05$. The z-score for kurtosis falls within the desired range, but the z-score for skewness does not. Using $\alpha = 0.05$, the sample has passed the normality assumption for kurtosis, yet failed the normality assumption for skewness. Therefore, either the sample must be modified and rechecked or you must use a nonparametric statistical test.

2.5   COMPUTING THE KOLMOGOROV–SMIRNOV ONE-SAMPLE TEST

The Kolmogorov–Smirnov one-sample test is a procedure to examine the agreement between two sets of values. For our purposes, the two sets of values compared are an observed frequency distribution based on a randomly collected sample and an empirical frequency distribution based on the sample's population. Furthermore, the observed sample is examined for normality when the empirical frequency distribution is based on a normal distribution.

The Kolmogorov–Smirnov one-sample test compares two cumulative frequency distributions. A cumulative frequency distribution is useful for finding the number of observations above or below a particular value in a data sample. It is calculated by taking a given frequency and adding all the preceding frequencies

in the list. In other words, it is like making a running total of the frequencies in a distribution. Creating cumulative frequency distributions of the observed and empirical frequency distributions allow us to find the point at which these two distributions show the largest divergence. Then, the test uses the largest divergence to identify a two-tailed probability estimate p to determine if the samples are statistically similar or different.

To perform the Kolmogorov–Smirnov one-sample test, we begin by determining the relative empirical frequency distribution $\hat{f}_{x_i}$ based on the observed sample. This relative empirical frequency distribution will approximate a normal distribution since we are examining our observed values for sample normality. First, calculate the observed frequency distribution's midpoint M and standard deviation s. The midpoint and standard deviation are found using Formula 2.11 and Formula 2.12:

$$M = (x_{max} + x_{min}) \div 2 \tag{2.11}$$

where x_{max} is the largest value in the sample and x_{min} is the smallest value in the sample, and

$$s = \sqrt{\frac{\sum (f_i x_i^2) - \dfrac{\left(\sum f_i x_i\right)^2}{n}}{n-1}} \tag{2.12}$$

where x_i is a given value in the observed sample, f_i is the frequency of a given value in the observed sample, and n is the number of values in the observed sample.

Next, use the midpoint and standard deviation to calculate the z-scores (see Formula 2.13) for the sample values x_i,

$$z = \left| \frac{x_i - M}{s} \right| \tag{2.13}$$

Use those z-scores and Table B.1 in Appendix B to determine the probability associated with each sample value, $\hat{p}_{x_i}$. These p-values are the relative frequencies of the empirical frequency distribution $\hat{f}_r$.

Now, we find the relative values of the observed frequency distribution f_r. Use Formula 2.14:

$$f_r = \frac{f_i}{n} \tag{2.14}$$

where f_i is the frequency of a given value in the observed sample and n is the number of values in the observed sample.

Since the Kolmogorov–Smirnov test uses cumulative frequency distributions, both the relative empirical frequency distribution and relative observed frequency distribution must be converted into cumulative frequency distributions $\hat{F}_{x_i}$ and S_{x_i}, respectively. Use Formula 2.15 and Formula 2.16 to find the absolute value divergence $\tilde{D}$ and D between the cumulative frequency distributions:

$$\tilde{D} = \left| \hat{F}_{x_i} - S_{x_i} \right| \tag{2.15}$$

$$D = \left| \hat{F}_{x_i} - S_{x_{i-1}} \right| \tag{2.16}$$

Use the largest divergence with Formula 2.17 to calculate the Kolmogorov–Smirnov test statistic Z:

$$Z = \sqrt{n} \max\left(|D|, |\tilde{D}|\right) \tag{2.17}$$

Then, use the Kolmogorov–Smirnov test statistic Z and the Smirnov (1948) formula (see Formula 2.18, Formula 2.19, Formula 2.20, Formula 2.21, Formula 2.22, and Formula 2.23) to find the two-tailed probability estimate p:

$$\text{if } 0 \leq Z < 0.27, \text{ then } p = 1 \tag{2.18}$$

$$\text{if } 0.27 \leq Z < 1, \text{ then } p = 1 - \frac{2.506628}{Z}(Q + Q^9 + Q^{25}) \tag{2.19}$$

where

$$Q = e^{-1.233701Z^{-2}} \tag{2.20}$$

$$\text{if } 1 \leq Z < 3.1, \text{ then } p = 2(Q - Q^4 + Q^9 - Q^{16}) \tag{2.21}$$

where

$$Q = e^{-2Z^2} \tag{2.22}$$

$$\text{if } Z \geq 3.1, \text{ then } p = 0 \tag{2.23}$$

A p-value that exceeds the level of risk associated with the null hypothesis indicates that the observed sample approximates the empirical sample. Since our empirical distributions approximated a normal distribution, we can state that our observed sample is sufficiently normal for parametric statistics. Conversely, a p-value that is smaller than the level of risk indicates an observed sample that is not sufficiently normal for parametric statistics. The nonparametric statistical tests in this book are useful if a sample lacks normality.

2.5.1   Sample Kolmogorov–Smirnov One-Sample Test

A department store has decided to evaluate customer satisfaction. As part of a pilot study, the store provides customers with a survey to rate employee friendliness. The survey uses a scale of 1–10 and its developer indicates that the scores should conform to a normal distribution. Use the Kolmogorov–Smirnov one-sample test to decide if the sample of customers surveyed responded with scores approximately matching a normal distribution. The survey results are shown in Table 2.5.

TABLE 2.5

Survey results

7	3	3	6
4	4	4	5
5	5	8	9
5	5	5	7
6	8	6	2

2.5.1.1   State the Null and Research Hypotheses   The null hypothesis states that the observed sample has an approximately normal distribution. The research hypothesis states that the observed sample does not approximately resemble a normal distribution.

The null hypothesis is

H_O: There is no difference between the observed distribution of survey scores and a normally distributed empirical sample.

The research hypothesis is

H_A: There is a difference between the observed distribution of survey scores and a normally distributed empirical sample.

2.5.1.2   Set the Level of Risk (or the Level of Significance) Associated with the Null Hypothesis   The level of risk, also called an alpha (α), is frequently set at 0.05. We will use an $\alpha = 0.05$ in our example. In other words, there is a 95% chance that any observed statistical difference will be real and not due to chance.

2.5.1.3   Choose the Appropriate Test Statistic   We are seeking to compare our observed sample against a normally distributed empirical sample. The Kolmogorov–Smirnov one-sample test will provide this comparison.

2.5.1.4   Compute the Test Statistic   First, determine the midpoint and standard deviation for the observed sample. Table 2.6 helps to manage the summations for this process.

TABLE 2.6

Survey score	Score frequency		
x_i	f_i	$f_i x_i$	$f_i x_i^2$
1	0	0	0
2	1	2	4
3	2	6	18
4	3	12	48
5	6	30	150
6	3	18	108
7	2	14	98
8	2	16	128
9	1	9	81
10	0	0	0
	$n = 20$	$\sum f_i x_i = 107$	$\sum f_i x_i^2 = 635$

Use Formula 2.11 to find the midpoint:

$$M = (x_{max} + x_{min}) \div 2$$
$$= (9 + 2) \div 2$$
$$M = 5.5$$

Then, use Formula 2.12 to find the standard deviation:

$$s = \sqrt{\frac{\sum (f_i x_i^2) - \frac{\left(\sum f_i x_i\right)^2}{n}}{n-1}}$$
$$= \sqrt{\frac{635 - \frac{107^2}{20}}{20 - 1}}$$
$$s = 1.81$$

Now, determine the z-scores, empirical relative frequencies, and observed relative frequencies for each score value (see Table 2.7).

TABLE 2.7

Survey score	Score frequency			Empirical frequency	Observed frequency
x_i	f_i	z-score	$\hat{p}_{x_i}$	$\hat{f}_r$	f_r
1	0	2.49	0.0064	0.006	0.000
2	1	1.93	0.0266	0.020	0.050
3	2	1.38	0.0838	0.064	0.100
4	**3**	**0.83**	**0.2033**	**0.140**	**0.150**
5	6	0.28	0.3897	0.250	0.300
6	3	0.28	0.3897	0.250	0.150
7	2	0.83	0.2033	0.140	0.100
8	2	1.38	0.0838	0.064	0.100
9	1	1.93	0.0266	0.020	0.050
10	0	2.49	0.0064	0.006	0.000

We will provide a sample calculation for survey score $= 4$ as seen in Table 2.7. Use Formula 2.13 to calculate the z-scores:

$$z = \left| \frac{x_i - M}{s} \right|$$
$$= \left| \frac{4 - 5.5}{1.81} \right|$$
$$z = 0.83$$

Use each z-score and Table B.1 in Appendix B to determine the probability associated with the each value, $\hat{p}_{x_i}$:

$$\hat{p}_4 = 0.2033$$

To find the empirical frequency value $\hat{f}_r$ for each value, subtract its preceding value, $\hat{f}_{r-1}$, from the associated probability value $\hat{p}_{x_i}$. In other words,

$$\hat{f}_r = \hat{p}_{x_i} - \hat{f}_{r-1}$$

We establish our empirical frequency distribution beginning at the tail, $x_i = 1$, and work to the midpoint, $x_i = 5$:

$$\hat{f}_{r1} = \hat{p}_1 - \hat{f}_{r0} = 0.0064 - 0.000 = 0.006$$
$$\hat{f}_{r2} = \hat{p}_2 - \hat{f}_{r1} = 0.0266 - 0.006 = 0.020$$
$$\hat{f}_{r3} = \hat{p}_3 - \hat{f}_{r2} = 0.0838 - 0.020 = 0.064$$
$$\hat{f}_{r4} = \hat{p}_4 - \hat{f}_{r3} = 0.2033 - 0.064 = 0.140$$
$$\hat{f}_{r5} = \hat{p}_5 - \hat{f}_{r4} = 0.3897 - 0.140 = 0.250$$

Our empirical frequency distribution is based on a normal distribution, which is symmetrical. Therefore, we can complete our empirical frequency distribution by basing the remaining values on a symmetrical distribution. Those values are in Table 2.7.

Now, we find the values of the observed frequency distribution f_r with Formula 2.14. We provide a sample calculation with survey result $= 4$. That survey value occurs three times:

$$f_{r4} = \frac{f_{x_i=4}}{n} = \frac{3}{20}$$
$$f_r = 0.150$$

Next, we create cumulative frequency distributions using the empirical and observed frequency distributions. A cumulative frequency distribution is created by taking a frequency and adding all the preceding values. We demonstrate this in Table 2.8.

Now, we find the absolute value divergence $\tilde{D}$ and D between the cumulative frequency distributions. Use Formula 2.15 and Formula 2.16. See the sample calculation for survey score $= 4$ as seen in bold in Table 2.9.

$$\tilde{D}_4 = \left|\hat{F}_4 - S_4\right| = |0.230 - 0.300|$$
$$\tilde{D}_4 = 0.070$$

and

$$D_4 = \left|\hat{F}_4 - S_3\right| = |0.230 - 0.150|$$
$$D_4 = 0.080$$

TABLE 2.8

| Survey score | Relative frequency | | Cumulative frequency | |
	Empirical	Observed	Empirical	Observed
x_i	$\hat{f}_r$	f_r	$\hat{F}_{x_i}$	S_{x_i}
1	0.006	0.000	0.006	0.000
2	0.020	0.050	$0.020 + 0.006 = 0.026$	$0.050 + 0.000 = 0.050$
3	0.064	0.100	$0.064 + 0.026 = 0.090$	$0.100 + 0.050 = 0.150$
4	0.140	0.150	$0.140 + 0.090 = 0.230$	$0.150 + 0.150 = 0.300$
5	0.250	0.300	$0.250 + 0.230 = 0.480$	$0.300 + 0.300 = 0.600$
6	0.250	0.150	$0.250 + 0.480 = 0.730$	$0.150 + 0.600 = 0.750$
7	0.140	0.100	$0.140 + 0.730 = 0.870$	$0.100 + 0.750 = 0.850$
8	0.064	0.100	$0.064 + 0.870 = 0.934$	$0.100 + 0.850 = 0.950$
9	0.020	0.050	$0.020 + 0.934 = 0.954$	$0.050 + 0.950 = 1.000$
10	0.006	0.000	$0.006 + 0.954 = 0.960$	$0.000 + 1.000 = 1.000$

TABLE 2.9

| Survey score | Cumulative frequency | | Cumulative frequency | |
	Empirical	Observed	Divergence	
x_i	$\hat{F}_{x_i}$	S_{x_i}	$\tilde{D}$	D
1	0.006	0.000	0.006	
2	0.026	0.050	0.024	0.026
3	0.090	0.150	0.060	0.040
4	**0.230**	**0.300**	**0.070**	**0.080**
*5	0.480	0.600	0.120	*0.180
6	0.730	0.750	0.020	0.130
7	0.870	0.850	0.020	0.120
8	0.934	0.950	0.016	0.084
9	0.954	1.000	0.046	0.004
10	0.960	1.000	0.040	0.040

To find the test statistic Z, use the largest value from $\tilde{D}$ and D in Formula 2.17. Table 2.9 has an asterisk next to the largest divergence. That value is located at survey value $= 5$. It is $\max(|D|, |\tilde{D}|) = 0.180$:

$$Z = \sqrt{n} \max(|D|, |\tilde{D}|)$$
$$= \sqrt{20}(0.180)$$
$$Z = 0.805$$

2.5.1.5 Determine the p-Value Associated with the Test Statistic The Kolmogorov–Smirnov test statistic Z and the Smirnov (1948) formula (see Formula 2.18, Formula 2.19, Formula 2.20, Formula 2.21, Formula 2.22, and Formula 2.23) are used to find the two-tailed probability estimate p. Since $0.27 \leq Z < 1$, we use Formula 2.19 and Formula 2.20:

$$Q = e^{-1.233701Z^{-2}}$$
$$= e^{-(1.233701)(0.805)^{-2}}$$
$$Q = 0.149$$

and

$$p = 1 - \frac{2.506628}{Z}(Q + Q^9 + Q^{25})$$
$$= 1 - \frac{2.506628}{0.805}(0.149 + 0.149^9 + 0.149^{25})$$
$$p = 0.536$$

2.5.1.6 Compare the p-Value with the Level of Risk (or the Level of Significance) Associated with the Null Hypothesis The critical value for rejecting the null hypothesis is $\alpha = 0.05$ and the obtained p-value is $p = 0.536$. If the critical value is greater than the obtained value, we must reject the null hypothesis. If the critical value is less than the obtained p-value, we must not reject the null hypothesis. Since the critical value is less than the obtained value ($0.05 < 0.536$), we do not reject the null hypothesis.

2.5.1.7 Interpret the Results We did not reject the null hypothesis, suggesting the customers' survey ratings of employee friendliness sufficiently resembled a normal distribution. This means that a parametric statistical procedure may be used with this sample.

2.5.1.8 Reporting the Results When reporting the results from the Kolmogorov–Smirnov one-sample test, we include the test statistic (D), the degrees of freedom (which equals the sample size), and the p-value in terms of the level of risk α. Based on our analysis, the sample of customers is approximately normal, where $D_{(20)} = 0.180$, $p > 0.05$.

2.5.2 Performing the Kolmogorov–Smirnov One-Sample Test Using SPSS

We will analyze the data from the example earlier using SPSS.

2.5.2.1 Define Your Variables First, click the "Variable View" tab at the bottom of your screen. Then, type the names of your variables in the "Name" column. As shown in Figure 2.13, the variable is called "Survey."

FIGURE 2.13

2.5.2.2   Type in Your Values   Click the "Data View" tab at the bottom of your screen. Type your sample values in the "Survey" column as shown in Figure 2.14.

FIGURE 2.14

2.5.2.3   Analyze Your Data   As shown in Figure 2.15, use the pull-down menus to choose "Analyze," "Nonparametric Tests," "Legacy Dialogs," and "1-Sample K-S . . ."

Use the arrow button to place your variable with your data values in the box labeled "Test Variable List:" as shown in Figure 2.16. Finally, click "OK" to perform the analysis.

2.5.2.4   Interpret the Results from the SPSS Output Window   SPSS Output 2.2 provides the most extreme difference ($D = 0.176$), Kolmogorov–Smirnov Z-test statistic ($Z = 0.789$), and the significance ($p = 0.562$). Based on the results from SPSS, the p-value exceeds the level of risk associated with the null hypothesis ($\alpha = 0.05$). Therefore, we do not reject the null hypothesis. In other words, the sample distribution is sufficiently normal.
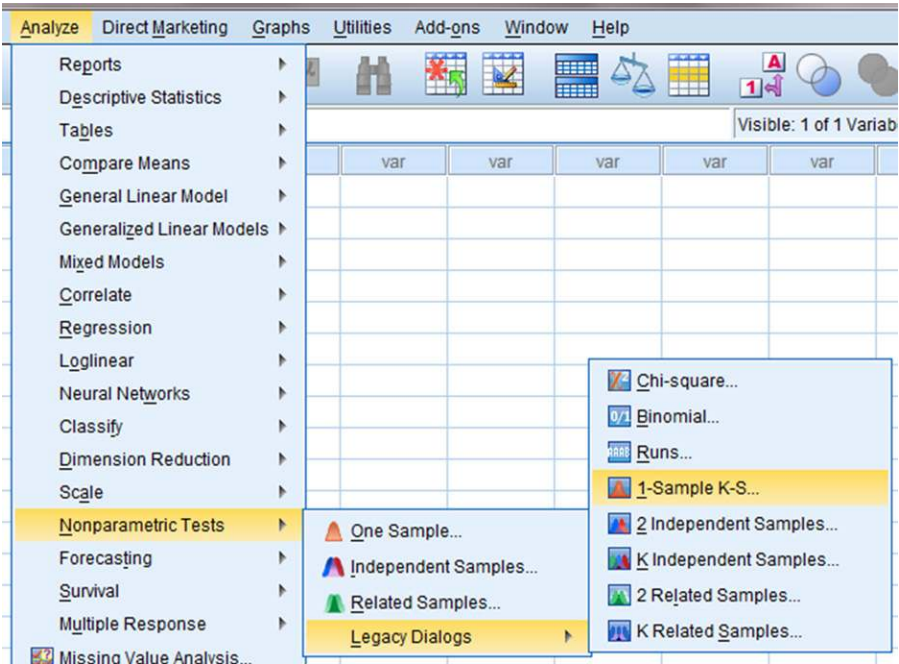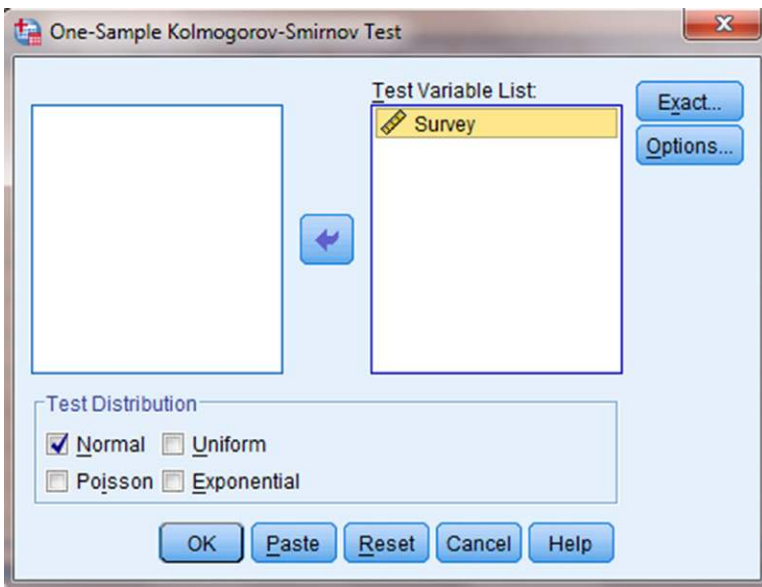
FIGURE 2.15

FIGURE 2.16

One-Sample Kolmogorov-Smirnov Test

		Survey
N		20
Normal Parameters[a,b]	Mean	5.3500
	Std. Deviation	1.81442
Most Extreme Differences	Absolute	.176
	Positive	.176
	Negative	-.124
Kolmogorov-Smirnov Z		.789
Asymp. Sig. (2-tailed)		.562

a. Test distribution is Normal.

b. Calculated from data.

SPSS OUTPUT 2.2

On an added note, differences between the values from the sample problem earlier and the SPSS output are likely due to value precision and computational round off errors.

2.6   SUMMARY

Parametric statistical tests, such as the *t*-test and one-way analysis of variance, are based on particular assumptions or parameters. Therefore, it is important that you examine collected data for its approximation to a normal distribution. Upon doing that, you can consider whether you will use a parametric or nonparametric test for analyzing your data.

In this chapter, we presented three quantitative measures of sample normality. First, we described how to examine a sample's kurtosis and skewness. Then, we described how to perform and interpret a Kolmogorov–Smirnov one-sample test. In the following chapters, we will describe several nonparametric procedures for analyzing data samples that do not meet the assumptions needed for parametric statistical tests. In the chapter that follows, we will begin by describing a test for comparing two unrelated samples.

2.7   PRACTICE QUESTIONS

1. The values in Table 2.10 are a sample of reading-level score for a 9th-grade class. They are measured on a ratio scale. Examine the sample's skewness and kurtosis for normality for $\alpha = 0.05$. Report your findings.

2. Using a Kolmogorov–Smirnov one-sample test, examine the sample of values from Table 2.10. Report your findings.

TABLE 2.10

Ninth-grade reading-level score

8.10	8.20	8.20	8.70	8.70	8.80	8.80	8.90	8.90	8.90
9.20	9.20	9.20	9.30	9.30	9.30	9.40	9.40	9.40	9.40
9.50	9.50	9.50	9.50	9.60	9.60	9.60	9.70	9.70	9.90

2.8 SOLUTIONS TO PRACTICE QUESTIONS

1. SPSS returned the following values:

   skewness $= -0.904$

   standard error of the skewness $= 0.427$

   kurtosis $= 0.188$

   standard error of the kurtosis $= 0.833$

   The computed z-scores are as follows:

   $$z_{S_k} = -2.117$$

   and

   $$z_K = 0.226$$

   At $\alpha = 0.05$, the sample's skewness fails the normality test, while the kurtosis passes the normality test. Based on our standard of $\alpha = 0.05$, this sample of reading levels for 9th-grade students is not sufficiently normal.

2. SPSS Output 2.3 shows the results from the Kolmogorov–Smirnov one-sample test.
   Kolmogorov–Smirnov obtained value $= 1.007$
   Two-Tailed significance $= 0.263$

One-Sample Kolmogorov-Smirnov Test

		Scores
N		30
Normal Parameters[a,b]	Mean	9.1800
	Std. Deviation	.46639
Most Extreme Differences	Absolute	.184
	Positive	.099
	Negative	-.184
Kolmogorov-Smirnov Z		1.007
Asymp. Sig. (2-tailed)		.263

a. Test distribution is Normal.
b. Calculated from data.

SPSS OUTPUT 2.3

According to the Kolmogorov–Smirnov one-sample test with $\alpha = 0.05$, this sample of reading levels for 9th-grade students is sufficiently normal.

COMPARING TWO RELATED SAMPLES: THE WILCOXON SIGNED RANK AND THE SIGN TEST

3.1 OBJECTIVES

In this chapter, you will learn the following items:

- How to compute the Wilcoxon signed rank test.
- How to perform the Wilcoxon signed rank test using SPSS®.
- How to construct a median confidence interval based on the Wilcoxon signed rank test for matched pairs.
- How to compute the sign test.
- How to perform the sign test using SPSS.

3.2 INTRODUCTION

Imagine that you give an attitude test to a small group of people. After you deliver some type of treatment, say, a daily vitamin C supplement for several weeks, you give that same group of people another attitude test. Finally, you compare the two measures of attitude to see if there is any type of difference between the two sets of scores.

The two sets of test scores in the previous scenario are related or paired. This is because each person was tested twice. In other words, each test score in one group of scores has another test score counterpart. The Wilcoxon signed rank test and the sign test are nonparametric statistical procedures for comparing two samples that are paired or related. The parametric equivalent to these tests goes by names such as the Student's *t*-test, *t*-test for matched pairs, *t*-test for paired samples, or *t*-test for dependent samples.

In this chapter, we will describe how to perform and interpret a Wilcoxon signed rank test and a sign test, using both small samples and large samples. In addition, we demonstrate the procedures for performing both tests using SPSS. Finally, we offer varied examples of these nonparametric statistics from the literature.

3.3   COMPUTING THE WILCOXON SIGNED RANK TEST STATISTIC

The formula for computing the Wilcoxon T for small samples is shown in Formula 3.1. The signed ranks are the values that are used to compute the positive and negative values in the formula:

$$T = \text{smaller of } \Sigma R_+ \text{ and } \Sigma R_- \tag{3.1}$$

where ΣR_+ is the sum of the ranks with positive differences and ΣR_- is the sum of the ranks with negative differences.

After the T statistic is computed, it must be examined for significance. We may use a table of critical values (see Table B.3 in Appendix B). However, if the numbers of pairs n exceeds those available from the table, then a large sample approximation may be performed. For large samples, compute a z-score and use a table with the normal distribution (see Table B.1 in Appendix B) to obtain a critical region of z-scores. Formula 3.2, Formula 3.3, and Formula 3.4 are used to find the z-score of a Wilcoxon signed rank test for large samples:

$$\bar{x}_T = \frac{n(n+1)}{4} \tag{3.2}$$

where $\bar{x}_T$ is the mean and n is the number of matched pairs included in the analysis,

$$s_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{3.3}$$

where s_T is the standard deviation,

$$z^* = \frac{T - \bar{x}_T}{s_T} \tag{3.4}$$

where z^* is the z-score for an approximation of the data to the normal distribution and T is the T statistic.

At this point, the analysis is limited to identifying the presence or absence of a significant difference between the groups and does not describe the strength of the treatment. We can consider the effect size (ES) to determine the degree of association between the groups. We use Formula 3.5 to calculate the ES:

$$ES = \frac{|z|}{\sqrt{n}} \tag{3.5}$$

where $|z|$ is the absolute value of the z-score and n is the number of matched pairs included in the analysis.

The ES ranges from 0 to 1. Cohen (1988) defined the conventions for ES as *small* $= 0.10$, *medium* $= 0.30$, and *large* $= 0.50$. (Correlation coefficient and ES are both measures of association. See Chapter 7 concerning correlation for more information on Cohen's assignment of ES's relative strength.)

3.3.1 Sample Wilcoxon Signed Rank Test (Small Data Samples)

The counseling staff of Clear Creek County School District has implemented a new program this year to reduce bullying in their elementary schools. The school district does not know if the new program resulted in improvement or deterioration. In order to evaluate the program's effectiveness, the school district has decided to compare the percentage of successful interventions last year before the program began with the percentage of successful interventions this year with the program in place. In Table 3.1, the 12 elementary school counselors, or participants, reported the percentage of successful interventions last year and the percentage this year.

TABLE 3.1

Participants	Percentage of successful interventions	
	Last year	This year
1	31	31
2	14	14
3	53	50
4	18	30
5	21	28
6	44	48
7	12	35
8	36	32
9	22	23
10	29	34
11	17	27
12	40	42

The samples are relatively small, so we need a nonparametric procedure. Since we are comparing two related, or paired, samples, we will use the Wilcoxon signed rank test.

3.3.1.1 State the Null and Research Hypotheses   The null hypothesis states that the counselors reported no difference in the percentages last year and this year. The research hypothesis states that the counselors observed some differences between this year and last year. Our research hypothesis is a two-tailed, nondirectional hypothesis because it indicates a difference, but in no particular direction.

The null hypothesis is

H_O: $\mu_D = 0$

The research hypothesis is

H_A: $\mu_D \neq 0$

3.3.1.2   Set the Level of Risk (or the Level of Significance) Associated with the Null Hypothesis

The level of risk, also called an alpha (α), is frequently set at 0.05. We will use $\alpha = 0.05$ in our example. In other words, there is a 95% chance that any observed statistical difference will be real and not due to chance.

3.3.1.3   Choose the Appropriate Test Statistic

The data are obtained from 12 counselors, or participants, who are using a new program designed to reduce bullying among students in the elementary schools. The participants reported the percentage of successful interventions last year and the percentage this year. We are comparing last year's percentages with this year's percentages. Therefore, the data samples are related or paired. In addition, sample sizes are relatively small. Since we are comparing two related samples, we will use the Wilcoxon signed rank test.

3.3.1.4   Compute the Test Statistic

First, compute the difference between each sample pair. Then, rank the absolute value of those computed differences. Using this method, the differences of zero are ignored when ranking. We have done this in Table 3.2.

TABLE 3.2

| Participant | Percentage of successful interventions | | Difference | Rank | Sign |
	Last year	This year		Without zero	
1	31	31	0	Exclude	
2	14	14	0	Exclude	
3	53	50	−3	3	−
4	18	30	+12	9	+
5	21	28	+7	7	+
6	44	48	+4	4.5	+
7	12	35	+23	10	+
8	36	32	−4	4.5	−
9	22	23	+1	1	+
10	29	34	+5	6	+
11	17	27	+10	8	+
12	40	42	+2	2	+

Compute the sum of ranks with positive differences. Using Table 3.2, the ranks with positive differences are 9, 7, 4.5, 10, 1, 6, 8, and 2. When we add all of the ranks with positive difference we get $\Sigma R_+ = 47.5$.

Compute the sum of ranks with negative differences. The ranks with negative differences are 3 and 4.5. The sum of ranks with negative difference is $\Sigma R_- = 7.5$.

The obtained value is the smaller of the two rank sums. Therefore, the Wilcoxon is $T = 7.5$.

3.3.1.5 Determine the Value Needed for Rejection of the Null Hypothesis Using the Appropriate Table of Critical Values for the Particular Statistic   Since the sample sizes are small, we use Table B.3 in Appendix B, which lists the critical values for the Wilcoxon T. As noted earlier in Table 3.2, the two counselors with score differences of zero were discarded. This reduces our sample size to $n = 10$. In this case, we look for the critical value under the two-tailed test for $n = 10$ and $\alpha = 0.05$. Table B.3 returns a critical value for the Wilcoxon test of $T = 8$. An obtained value that is less than or equal to 8 will lead us to reject our null hypothesis.

3.3.1.6 Compare the Obtained Value with the Critical Value   The critical value for rejecting the null hypothesis is 8 and the obtained value is $T = 7.5$. If the critical value equals or exceeds the obtained value, we must reject the null hypothesis. If instead, the critical value is less than the obtained value, we must not reject the null hypothesis. Since the critical value exceeds the obtained value, we must reject the null hypothesis.

3.3.1.7 Interpret the Results   We rejected the null hypothesis, suggesting that a real difference exists between last year's percentages and this year's percentages. In addition, since the sum of the positive difference ranks (ΣR_+) was larger than the negative difference ranks (ΣR_-), the difference is positive, showing a positive impact of the program. Therefore, our analysis provides evidence that the new bullying program is providing positive benefits toward the improvement of student behavior as perceived by the school counselors.

3.3.1.8 Reporting the Results   When reporting the findings, include the T statistic, sample size, and p-value's relation to α. The directionality of the difference should be expressed using the sum of the positive difference ranks (ΣR_+) and sum of the negative difference ranks (ΣR_-).

For this example, the Wilcoxon signed rank test ($T = 7.5$, $n = 12$, $p < 0.05$) indicated that the percentage of successful interventions was significantly different. In addition, the sum of the positive difference ranks ($\Sigma R_+ = 47.5$) was larger than the sum of the negative difference ranks ($\Sigma R_- = 7.5$), showing a positive impact from the program. Therefore, our analysis provides evidence that the new bullying program is providing positive benefits toward the improvement of student behavior as perceived by the school counselors.

3.3.2 Confidence Interval for the Wilcoxon Signed Rank Test

The American Psychological Association (2001) has suggested that researchers report the *confidence interval* for research data. A confidence interval is an inference to a population in terms of an estimation of sampling error. More specifically, it provides a range of values that fall within the population with a level of confidence of $100(1 - \alpha)\%$.

A median confidence interval can be constructed based on the Wilcoxon signed rank test for matched pairs. In order to create this confidence interval, all of the possible matched pairs (X_i, X_j) are used to compute the differences $D_i = X_i - X_j$. Then, compute all of the averages u_{ij} of two difference scores using Formula 3.6. There will be a total of $[n(n-1)/2] + n$ averages.

$$u_{ij} = (D_i + D_j)/2 \quad 1 \le i \le j \le n \tag{3.6}$$

We will perform a 95% confidence interval using the sample Wilcoxon signed rank test with a small data sample (as stated earlier). Table 3.1 provides the values for obtaining our confidence interval. We begin by using Formula 3.6 to compute all of the averages u_{ij} of two difference scores. For example,

$$u_{11} = (D_1 + D_1)/2 = (-3 + -3)/2$$

$$u_{11} = -3$$

$$u_{12} = (D_1 + D_2)/2 = (-3 + 12)/2$$

$$u_{12} = 4.5$$

$$u_{13} = (D_1 + D_3)/2 = (-3 + 7)/2$$

$$u_{13} = 2$$

Table 3.3 shows each value of u_{ij}.

TABLE 3.3

	−3	12	7	4	23	−4	1	5	10	2
−3	−3	4.5	2	0.5	10	−3.5	−1	1	3.5	−0.5
12		12	9.5	8	17.5	4	6.5	8.5	11	7
7			7	5.5	15	1.5	4	6	8.5	4.5
4				4	13.5	0	2.5	4.5	7	3
23					23	9.5	12	14	16.5	12.5
−4						−4	−1.5	0.5	3	−1
1							1	3	5.5	1.5
5								5	7.5	3.5
10									10	6
2										2

Next, arrange all of the averages in order from smallest to largest. We have arranged all of the values for u_{ij} in Table 3.4.

The median of the ordered averages gives a point estimate of the population median difference. The median of this distribution is 4.5, which is the point estimate of the population.

Use Table B.3 in Appendix B to find the endpoints of the confidence interval. First, determine T from the table that corresponds with the sample size and desired

TABLE 3.4

1	−4.0	12	1.0	22	4.0	34	6.5	45	10.0
2	−3.5	13	1.5	23	4.0	35	7.0	46	11.0
3	−3.0	14	1.5	24	4.0	36	7.0	47	12.0
4	−1.5	15	2.0	25	4.5	37	7.0	48	12.0
5	−1.0	15	2.0	26	4.5	38	7.5	49	12.5
6	−1.0	16	2.5	27	4.5	39	8.0	50	13.5
7	−0.5	17	3.0	28	5.0	40	8.5	51	14.0
8	0.0	18	3.0	29	5.5	41	8.5	52	15.0
9	0.5	19	3.0	30	5.5	42	9.5	53	16.5
10	0.5	20	3.5	31	6.0	43	9.5	54	17.5
11	1.0	21	3.5	32	6.0	44	10.0	55	23.0

confidence such that $p = \alpha/2$. We seek to find a 95% confidence interval. For our example, $n = 10$ and $p = 0.05/2$. The table provides $T = 8$.

The endpoints of the confidence interval are the Kth smallest and the Kth largest values of u_{ij}, where $K = T + 1$. For our example, $K = 8 + 1 = 9$. The ninth value from the bottom is 0.5 and the ninth value from the top is 12.0. Based on these findings, it is estimated with 95% confident that the difference of successful interventions due to the new bullying programs lies between 0.5 and 12.0.

3.3.3 Sample Wilcoxon Signed Rank Test (Large Data Samples)

Hearing of Clear Creek School District's success with their antibullying program, Jonestown School District has implemented the program this year to reduce bullying in their own elementary schools. The Jonestown School District evaluates their program's effectiveness by comparing the percentage of successful interventions last year before the program began with the percentage of successful interventions this year with the program in place. In Table 3.5, the 25 elementary school counselors, or participants, reported the percentage of successful interventions last year and the percentage this year.

TABLE 3.5

Participant	Percentage of successful interventions	
	Last year	This year
1	53	50
2	18	43
3	21	28
4	44	48
5	12	35
6	36	32

(*Continued*)

TABLE 3.5 (*Continued*)

	Percentage of successful interventions	
Participant	Last year	This year
7	22	23
8	29	34
9	17	27
10	10	42
11	38	44
12	37	16
13	19	33
14	37	50
15	28	20
16	15	27
17	25	27
18	38	30
19	40	51
20	30	50
21	23	45
22	41	20
23	31	49
24	28	43
25	14	30

We will use the same nonparametric procedure to analyze the data. However, use a large sample ($n \geq 20$) approximation.

3.3.3.1 State the Null and Research Hypotheses
The null hypothesis states that the counselors reported no difference in the percentages last year and this year. The research hypothesis states that the counselors observed some differences between this year and last year. Our research hypothesis is a two-tailed, nondirectional hypothesis because it indicates a difference, but in no particular direction.

The null hypothesis is

H_O: $\mu_D = 0$

The research hypothesis is

H_A: $\mu_D \neq 0$

3.3.3.2 Set the Level of Risk (or the Level of Significance) Associated with the Null Hypothesis
The level of risk, also called an alpha (α), is frequently set at 0.05. We will use $\alpha = 0.05$ in our example. In other words, there is a 95% chance that any observed statistical difference will be real and not due to chance.

3.3.3.3 Choose the Appropriate Test Statistic   The data are obtained from 25 counselors, or participants, who are using a new program designed to reduce bullying among students in the elementary schools. The participants reported the percentage of successful interventions last year and the percentage this year. We are comparing last year's percentages with this year's percentages. Therefore, the data samples are related or paired. Since we are comparing two related samples, we will use the Wilcoxon signed rank test.

3.3.3.4 Compute the Test Statistic   First, compute the difference between each sample pair. Then, rank the absolute value of those computed differences. We have done this in Table 3.6.

TABLE 3.6

Participant	Percentage of successful interventions		Difference	Rank	Sign
	Last year	This year			
1	53	50	−3	3	−
2	18	43	+25	24	+
3	21	28	+7	8	+
4	44	48	+4	4.5	+
5	12	35	+23	23	+
6	36	32	−4	4.5	−
7	22	23	+1	1	+
8	29	34	+5	6	+
9	17	27	+10	11	+
10	10	42	+32	25	+
11	38	44	+6	7	+
12	37	16	−21	20.5	−
13	19	33	+14	15	+
14	37	50	+13	14	+
15	28	20	−8	9.5	−
16	15	27	+12	13	+
17	25	27	+2	2	+
18	38	30	−8	9.5	−
19	40	51	+11	12	+
20	30	50	+20	19	+
21	23	45	+22	22	+
22	41	20	−21	20.5	−
23	31	49	+18	18	+
24	28	43	+15	16	+
25	14	30	+16	17	+

Compute the sum of ranks with positive differences. Using Table 3.6, when we add all of the ranks with positive difference, we get $\Sigma R_+ = 257.5$.

Compute the sum of ranks with negative differences. The ranks with negative differences are 3, 4.5, 9.5, 9.5, 20.5, and 20.5. The sum of ranks with negative difference is $\Sigma R_- = 67.5$.

The obtained value is the smaller of these two rank sums. Thus, the Wilcoxon $T = 67.5$.

Since our sample size is larger than 20, we will approximate it to a normal distribution. Therefore, we will find a z-score for our data using a normal approximation. We must find the mean $\bar{x}_T$ and the standard deviation s_T for the data:

$$\bar{x}_T = \frac{n(n+1)}{4} = \frac{25(25+1)}{4}$$

$$\bar{x}_T = 162.5$$

and

$$s_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{25(25+1)(50+1)}{24}} = \sqrt{\frac{33,150}{24}}$$

$$s_T = 37.17$$

Next, we use the mean, standard deviation, and the T-test statistic to calculate a z-score. Remember, we are testing the hypothesis that there is no difference in ranks of percentages of successful interventions between last year and this year:

$$z^* = \frac{T - \bar{x}_T}{s_T} = \frac{67.5 - 162.5}{37.17}$$

$$z^* = -2.56$$

3.3.3.5   Determine the Value Needed for Rejection of the Null Hypothesis Using the Appropriate Table of Critical Values for the Particular Statistic
Table B.1 in Appendix B is used to establish the critical region of z-scores. For a two-tailed test with $\alpha = 0.05$, we must not reject the null hypothesis if $-1.96 \leq z^* \leq 1.96$.

3.3.3.6   Compare the Obtained Value to the Critical Value
We find that z^* is not within the critical region of the distribution, $-2.56 < -1.96$. Therefore, we reject the null hypothesis. This suggests a difference in the percentage of successful interventions after the program was implemented.

3.3.3.7   Interpret the Results
We rejected the null hypothesis, suggesting that a real difference exists between last year's percentages and this year's percentages. In addition, since the sum of the positive difference ranks (ΣR_+) was larger than the negative difference ranks (ΣR_-), the difference is positive, showing a positive impact of the program. Therefore, our analysis provides evidence that the new bullying program is providing positive benefits toward the improvement of student behavior as perceived by the school counselors.

At this point, the analysis is limited to identifying the presence or absence of a significant difference between the groups. In other words, the statistical test's level of significance does not describe the strength of the treatment. The American Psychological Association (2001), however, has called for a measure of the strength called the ES.

We can consider the ES for this large sample test to determine the degree of association between the groups. We use Formula 3.5 to calculate the ES. For the example, $|z| = 2.56$ and $n = 25$:

$$ES = \frac{|z|}{\sqrt{n}} = \frac{|-2.56|}{\sqrt{25}}$$

$$ES = 0.51$$

Our ES for the matched-pair samples is 0.51. This value indicates a high level of association between the percentage of successful interventions before and after the implementation of the new bullying program.

3.3.3.8   *Reporting the Results*

For this example, the Wilcoxon signed rank test ($T = 67.5$, $n = 25$, $p < 0.05$) indicated that the percentage of successful interventions was significantly different. In addition, the sum of the positive difference ranks ($\Sigma R_+ = 257.5$) was larger than the sum of the negative difference ranks ($\Sigma R_- = 67.5$), showing a positive impact from the program. Moreover, the ES for the matched-pair samples was 0.51. Therefore, our analysis provides evidence that the new bullying program is providing positive benefits toward the improvement of student behavior as perceived by the school counselors.

3.4   COMPUTING THE SIGN TEST

You can analyze related samples more efficiently by reducing values to dichotomous results ("yes" or "no") or ("+" or "−"). The sign test allows you to perform that analysis. Our procedure for performing the sign test is based on the method described by Gibbons and Chakraborti (2010).

We begin the procedure for performing a sign test by identifying whether each set from the related data samples demonstrates a positive difference, a negative difference, or no difference at all. Then, we find the sum of the positive differences n_p and the sum of negative differences n_n. Cases with no difference are ignored.

We perform the next part of the analysis based on the sum of differences. If $n_p + n_n = 0$, then the one-sided probability is $p = 0.5$. If $0 < n_p + n_n < 25$, then p is calculated recursively from the binomial probability function using Formula 3.7. Table B.9 in Appendix B includes several factorials to simplify computation:

$$P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot (1-p)^{n-X} \tag{3.7}$$

where $n = n_p + n_n$ and p is the probability of event occurrence.

If $n_p + n_n \geq 25$, we use Formula 3.8: