© Kitch/Age Fotostock America, Inc.

Chapter

15

Nonparametric Methods[1]

What kind of soda is that on your desk? Regardless of the particular brand or flavor, it is more likely than last year that it is a diet soda. According to a Dow Jones report (August 19, 2002), diet sodas represent 30% of the soft drink market, a 6.6% sales increase over the prior year (compared to a 3.1% increase in sales of regular soft drinks). Even so, diet soft drinks represent only 18.2% of the total U.S. carbonated soft drink market, according to John Sicher, editor and publisher of *Beverage Digest*. We can conduct hypothesis tests to determine people's preferences for one type of soft drink over another.

The hypothesis tests discussed so far in this text are called **parametric tests**. In those tests, we used the normal, *t*, chi-square, and *F* distributions to make tests about population parameters such as means, proportions, variances, and standard deviations. In doing so, we made some assumptions, such as the assumption that the population from which the sample was drawn was normally distributed. This chapter discusses a few **nonparametric tests**. These tests do not require the same kinds of assumptions, and hence, they are also called **distribution-free tests**.

Nonparametric tests have several advantages over parametric tests: They are easier to use and understand; they can be applied to situations in which parametric tests cannot be used; and they do not require that the population being sampled is normally distributed. However, a major problem with nonparametric tests is that they are less efficient than parametric tests. The sample size must be larger for a nonparametric test to have the same probability of committing the two types of errors.

Although there are a large number of nonparametric tests that can be applied to conduct tests of hypothesis, this chapter discusses only six of them: the sign test, the Wilcoxon signed-rank test, the Wilcoxon rank sum test, the Kruskal-Wallis test, the Spearman rho rank correlation coefficient test, and the runs test for randomness.

[1]Tables VIII to XII that are needed for this chapter are given at the end of this chapter. Tables IV and VI are in Appendix C of the book.

15.1 The Sign Test

The **sign test** is one of the easiest tests to apply to test hypotheses. It uses only plus and minus signs. The sign test can be used to perform the following types of tests:

1. To determine the preference for one product or item over another, or to determine whether one outcome occurs more often than another outcome in categorical data. For example, we may test whether or not people prefer one kind of soft drink over another kind.
2. To conduct a test for the median of a single population. For example, we may use this procedure to test whether the median rent paid by all tenants in a city is different from $1250.
3. To perform a test for the median of paired differences using data from two dependent samples. For example, we may use this procedure to test whether the median score on a standardized test increases after a preparatory course is taken.

Definition

Sign Test The *sign test* is used to make hypothesis tests about preferences, a single median, and the median of paired differences for two dependent populations. We use only plus and minus signs to perform these tests.

In the following subsections we discuss these tests for small and large samples.

15.1.1 Tests About Categorical Data

Data that are divided into different categories for identification purposes are called **categorical data**. For example, people's opinions about a certain issue—in favor, against, or no opinion—produce categorical data. This subsection discusses how to perform tests about such data using the sign test procedure. We discuss two situations in which such tests can be performed: the small-sample case and the large-sample case.

The Small-Sample Case

When we apply the sign test for categorical data, if the *sample size is 25 or less* (i.e., $n \leq 25$), we consider it a small sample. Table VIII: Critical Values of X for the Sign Test (that appears at the end of this chapter), is based on the binomial probability distribution. This table gives the critical values of the test statistic for the sign test when $n \leq 25$ using the binomial probability distribution.

The sign test can be used to test whether or not customers prefer one brand of a product over another brand of the same type of product. For example, we can test whether customers have a higher preference for Coke or for Pepsi. This procedure can also be used to test whether people prefer one of two alternatives over the other. For example, given a choice, do people prefer to live in New York City or in Los Angeles?

■ EXAMPLE 15–1

Performing sign test with categorical data: small sample.

The Top Taste Water Company produces and distributes Top Taste bottled water. The company wants to determine whether customers have a higher preference for its bottled water than for its main competitor, Spring Hill bottled water. The Top Taste Water Company hired a statistician to conduct this study. The statistician selected a random sample of 10 people and asked each of them to taste one sample of each of the two brands of water. The customers did not know the brand of each water sample. Also, the order in which each person tasted the two brands of water was determined randomly. Each person was asked to indicate which of the two samples of water he or she preferred. The following table shows the preferences of these 10 individuals.

Person	Brand Preferred
1	Spring Hill
2	Top Taste
3	Top Taste
4	Neither
5	Top Taste
6	Spring Hill
7	Spring Hill
8	Top Taste
9	Top Taste
10	Top Taste

Based on these results, can the statistician conclude that people prefer one brand of bottled water over the other? Use the significance level of 5%.

Solution We use the same five steps to perform this test of hypothesis that we used in earlier chapters.

Step 1. *State the null and alternative hypotheses.*

If we assume that people do not prefer either brand of water over the other, we would expect about 50% of the people (among those who show a preference) to indicate a preference for Top Taste water and the other 50% to indicate a preference for Spring Hill water. Let p be the proportion of all people who prefer Top Taste bottled water. The two hypotheses are as follows:

$$H_0\text{: } p = .50 \quad \text{(People do not prefer either of the two brands of water)}$$

$$H_1\text{: } p \neq .50 \quad \text{(People prefer one brand of water over the other)}$$

The null hypothesis states that 50% of the people prefer Top Taste water over Spring Hill water (and, hence, the other 50% prefer Spring Hill water). Note that we do not consider people who have no preference, and we drop them from the sample. If we fail to reject H_0, we will conclude that people do not prefer one brand over the other. However, if we reject H_0, we will conclude that the percentage of people who prefer Top Taste water over Spring Hill water is different from 50%. Thus, the conclusion will be that people prefer one brand over the other.

Step 2. *Select the distribution to use.*

We use the binomial probability distribution to make the test. Note that here there is only one sample and each member of the sample is asked to indicate a preference if he or she has one. We drop the members who do not indicate a preference and then compare the preferences of the remaining members. Also note that there are three outcomes for each person: (1) prefers Top Taste water, (2) prefers Spring Hill water, or (3) has no preference. We are to compare the two outcomes with preferences and determine whether more people belong to one of these two outcomes than to the other. All such tests of preferences are conducted by using the binomial probability distribution. If we assume that H_0 is true, then the number of people who indicate a preference for Top Taste bottled water (the number of successes) follows the binomial distribution, with $p = .50$.

Step 3. *Determine the rejection and nonrejection regions.*

Note that 10 people were selected to taste the two brands of water and indicate their preferences. However, one of these individuals stated no preference. Hence, only 9 of the 10 people have indicated a preference for one or the other of the two brands of bottled water. To conduct the test, the person who has shown no preference is dropped from the sample. Thus, the *true sample size* is 9; that is, $n = 9$.

The significance level for the test is .05. Let X be the number of people in the sample of 9 who prefer Top Taste bottled water. Here X is called the *test statistic*. To establish a decision rule, we find the critical values of X from Table VIII for $n = 9$. From that table, for $n = 9$ and $\alpha = .05$ for a two-tailed test, the critical values of X are 1 and 8. Note that in a two-tailed test we read both the lower and the upper critical values, which are shown in Figure 15.1.

Thus, we will reject the null hypothesis if either fewer than two or more than seven people in nine indicate a preference for Top Taste bottled water.

Figure 15.1

0 or 1	2 to 7	8 or 9	X
Rejection region	Nonrejection region	Rejection region	

Critical Value(s) of *X* In a sign test for a small sample, the *critical value of X* is obtained from Table VIII. If the test is two-tailed, we read both the lower and the upper critical values from that table. However, we read only the lower critical value if the test is left-tailed, and only the upper critical value if the test is right-tailed. Also note that which column we use to obtain this critical value depends on the given significance level and on whether the test is two-tailed or one-tailed.

Step 4. *Calculate the value of the test statistic.*

To record the results of the experiment, we mark a plus sign for each person who prefers Top Taste bottled water, a minus sign for each person who prefers Spring Hill bottled water, and a zero for the one who indicates no preference. This listing is shown in Table 15.1.

Table 15.1

Person	Brand Preferred	Sign
1	Spring Hill	−
2	Top Taste	+
3	Top Taste	+
4	Neither	0
5	Top Taste	+
6	Spring Hill	−
7	Spring Hill	−
8	Top Taste	+
9	Top Taste	+
10	Top Taste	+

Now, we count the number of plus signs (the sign that belongs to Top Taste bottled water because p in H_0 refers to Top Taste water). There are six plus signs, indicating that six of the nine people in the sample stated a preference for Top Taste bottled water. Note that the sample size is 9, not 10, because we drop the person with zero sign. Thus,

$$\text{Observed value of } X = 6$$

Observed Value of *X* The *observed value of X* is given by the number of signs that belong to the category whose proportion we are testing for.

Step 5. *Make a decision.*

Because the observed value of $X = 6$ falls in the nonrejection region (see Figure 15.1), we fail to reject H_0. Hence, we conclude that our sample does not indicate that people prefer either of these two brands of bottled water over the other.

Note that it does not matter which outcome p refers to. If we assume that p is the proportion of people who prefer Spring Hill water, then X will denote the number of people in a sample of n who prefer Spring Hill water. The observed value of X this time would be 3, which is the number of minus signs in Table 15.1. From Figure 15.1, $X = 3$ also falls in the nonrejection region. Hence, we again fail to reject the null hypothesis. ■

The Large-Sample Case

If we are testing a hypothesis about preference for categorical data and $n > 25$, we can use the normal probability distribution as an approximation to the binomial probability distribution.

The Large-Sample Case If $n > 25$, the normal distribution can be used as an approximation to the binomial probability distribution to perform a test of hypothesis about the preference for categorical data. The *observed value* of the test statistic z, in this case, is calculated as

$$z = \frac{(X \pm .5) - \mu}{\sigma}.$$

where X is the number of units in the sample that belong to the outcome referring to p. We either add .5 to X or subtract .5 from X to correct for continuity (see Section 6.7 of Chapter 6). We will add .5 to X if the value of X is less than or equal to $n/2$, and we will subtract .5 from X if the value of X is greater than $n/2$. The values of the mean and standard deviation are calculated as

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

Example 15–2 illustrates the procedure for the large-sample case.

■ EXAMPLE 15–2

Performing sign test with categorical data: large sample.

A developer is interested in building a shopping mall adjacent to a residential area. Before granting or denying permission to build such a mall, the town council took a random sample of 75 adults from adjacent areas and asked them whether they favor or oppose construction of this mall. Of these 75 adults, 40 opposed construction of the mall, 30 favored it, and 5 had no opinion. Can you conclude that the number of adults in this area who oppose construction of the mall is higher than the number who favor it? Use $\alpha = .01$.

Solution Again, each adult in the sample has to pick one of three choices: oppose, favor, or have no opinion. And we are to compare two outcomes—oppose and favor—to find out whether more adults belong to the outcome indicated by *oppose*. We can use the sign test here. To do so, we drop the subjects who have no opinion—that is, the adults who belong to the outcome that is not being compared. In our example, five adults have no opinion. Thus, we drop these adults from our sample and use the sample size of $n = 70$ for the purposes of this test. Let p be the proportion of adults who oppose construction of the mall and q be the proportion who favor it. We apply the five steps to make this test.

Step 1. *State the null and alternative hypotheses.*

H_0: $p = .50$ and $q = .50$ (The two proportions are equal)

H_1: $p > .50$ or $p > q$ (The proportion of adults who oppose the mall is greater than the proportion who favor it)

The null hypothesis here states that the proportion of adults who oppose and the proportion who favor construction of the mall are both .50, which means that $p = .50$ and $q = .50$. The alternative hypothesis is that $p > q$, which means that more adults oppose construction of the mall than favor it. Note that H_1 states that $p > .50$ and $q < .50$. In other words, of those who have an opinion, more than 50% oppose and less than 50% favor construction of the shopping mall.

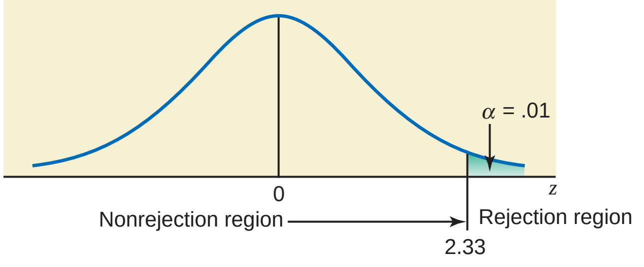
Step 2. *Select the distribution to use.*

As explained earlier, we will use the sign test to perform this test. Although 75 adults were asked their opinion, only 70 of them offered it and 5 did not. Hence, our sample size is 70; that is, $n = 70$. Because it is a large sample ($n > 25$), we can use the normal approximation to perform the test.

Step 3. *Determine the rejection and nonrejection regions.*

Because H_1 states that $p > .50$, our test is right-tailed. Also, $\alpha = .01$. From Table IV (the standard normal distribution table) in Appendix C, the z value for $1.0 - .01 = .9900$ area to the left is approximately 2.33. Thus, the decision rule is that we will not reject H_0 if $z < 2.33$ and we will reject H_0 if $z \geq 2.33$. Thus, the nonrejection region lies to the left of $z = 2.33$ and the rejection region to the right of $z = 2.33$, as shown in Figure 15.2.

Figure 15.2



Step 4. *Calculate the value of the test statistic.*

Assuming that the null hypothesis is true, we expect (about) half of the adults in the population to oppose construction of the mall and the other half to favor it. Thus, we expect $p = .50$ and $q = .50$. Note that we do not count the people who have no opinion. The mean and standard deviation of the binomial distribution are

$$\mu = np = 70(.50) = 35$$

$$\sigma = \sqrt{npq} = \sqrt{70(.50)(.50)} = \sqrt{17.5} = 4.18330013$$

In this example, p refers to the proportion of adults who oppose construction of the mall. Hence, X refers to the number in 70 who oppose the mall. Thus,

$$X = 40 \quad \text{and} \quad \frac{n}{2} = \frac{70}{2} = 35$$

Because X is greater than $n/2$, the observed value of the test statistic z is

$$z = \frac{(X - .5) - \mu}{\sigma} = \frac{(40 - .5) - 35}{4.18330013} = 1.08$$

Step 5. *Make a decision.*

Because the observed value of $z = 1.08$ is less than the critical value of $z = 2.33$, it falls in the nonrejection region. Hence, we do not reject H_0. Consequently, we conclude that the number of adults who oppose construction of the mall is not higher than the number who favor its construction. ■

15.1.2 Tests About the Median of a Single Population

The sign test can be used to test a hypothesis about a population median. Recall from Chapter 3 that the median is the value that divides a ranked data set into two equal parts. For example, if the median age of students in a class is 24, half of the students are younger than 24 and half are older. This section discusses how to make a test of hypothesis about the median of a population.

The Small-Sample Case

If $n \leq 25$, we use the binomial probability distribution to test a hypothesis about the median of a population. The procedure used to conduct such a test is very similar to the one explained in Example 15–1.

■ EXAMPLE 15–3

Performing sign test about a population median: small sample.

A real estate agent claims that the median price of homes in a small midwest city is \$137,000. A sample of 10 houses selected by a statistician produced the following data on their prices.

Home	1	2	3	4	5	6	7	8	9	10
Price (\$)	147,500	123,600	139,000	168,200	129,450	132,400	156,400	188,210	198,425	215,300

Using the 5% significance level, can you conclude that the median price of homes in this city is different from \$137,000?

Solution Using the given data, we prepare Table 15.2, which contains a sign row. In this row, we assign a positive sign to each price that is above the claimed median price of \$137,000 and a negative sign to each price that is below the claimed median price.

In Table 15.2, there are seven plus signs, indicating that the prices of seven houses are higher than the claimed median price of \$137,000, and there are three minus signs, showing that the prices of three homes are lower than the claimed median price. Note that if one or more values in a data set are equal to the median, then each of them is assigned a zero value and dropped from the sample. Next, we perform the following five steps to perform the test of hypothesis.

Table 15.2

Home	1	2	3	4	5	6	7	8	9	10
Sign	+	−	+	+	−	−	+	+	+	+

Step 1. *State the null and alternative hypotheses.*

$$H_0\text{: Median price} = \$137{,}000 \quad \text{(Real estate agent's claim is true)}$$

$$H_1\text{: Median price} \neq \$137{,}000 \quad \text{(Real estate agent's claim is false)}$$

Step 2. *Select the distribution to use.*

For a test of the median of a population, we employ the sign test procedure by using the binomial probability distribution if $n \leq 25$. Since in our example $n = 10$, which is less than 25, we use the binomial probability distribution to conduct the test.

Step 3. *Determine the rejection and nonrejection regions.*

In our example, $n = 10$ and $\alpha = .05$. The test is two-tailed. Let X be the test statistic that represents the number of plus signs in Table 15.2. From Table VIII, the (lower and upper) critical values of X are 1 and 9. Using these critical values, Figure 15.3 shows the rejection and nonrejection regions. Thus, we will reject the null hypothesis if the observed value of X is either 1 or 0, or 9 or 10. Note that because X represents the number of plus signs in the sample, its lowest possible value is 0 and its highest possible value is 10.

Figure 15.3

0 or 1	2 to 8	9 or 10	
Rejection region	Nonrejection region	Rejection region	X

Step 4. *Calculate the value of the test statistic.*

The observed value of X is given by the number of plus signs in Table 15.2. Thus,

$$\text{Observed value of } X = 7$$

Observed Value of *X* When using the sign test to perform a test about a median, we can use either the number of positive signs or the number of negative signs as the observed value of X if the test is two-tailed. However, the *observed value of X* is equal to the larger of these two numbers (the number of positive and negative signs) if the test is right-tailed, and equal to the smaller of these two numbers if the test is left-tailed.

Step 5. *Make a decision.*

The observed value of $X = 7$ falls in the nonrejection region in Figure 15.3. Hence, we do not reject H_0 and conclude that the median price of homes in this city is not different from \$137,000. ■

The Large-Sample Case

For a test of the median of a single population, we can use the normal approximation to the binomial probability distribution when $n > 25$. The observed value of z, in this case, is calculated as in a test of hypothesis about the preference for categorical data (see the rule described on page 635 and Example 15–2). Example 15–4 explains the procedure for such a test.

■ EXAMPLE 15–4

Performing sign test about a population median: large sample.

A long-distance phone company believes that the median phone bill (for long-distance calls) is at least \$70 for all the families in New Haven, Connecticut. A random sample of 90 families selected from New Haven showed that the phone bills of 51 of them were less than \$70 and those of 38 of them were more than \$70, and 1 family had a phone bill of exactly \$70. Using the 1% significance level, can you conclude that the company's claim is true?

Solution We use the usual five steps to test this hypothesis.

Step 1. *State the null and alternative hypotheses.*

The company's claim is that the median phone bill is at least \$70. Hence, the two hypotheses are as follows:

$$H_0\text{: Median} \geq \$70 \quad \text{(Company's claim is true)}$$
$$H_1\text{: Median} < \$70 \quad \text{(Company's claim is false)}$$

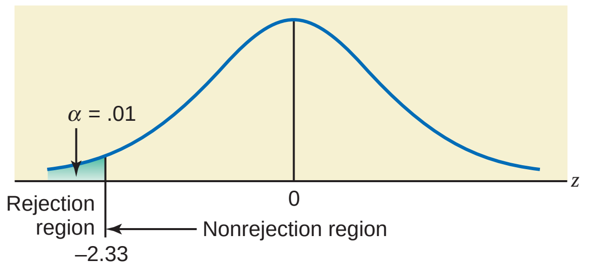
Step 2. *Select the distribution to use.*

This is a test about the median and $n > 25$. Hence, to conduct this test, we can use the normal distribution as an approximation to the binomial probability distribution.

Step 3. *Determine the rejection and nonrejection regions.*

The test is left-tailed and $\alpha = .01$. From Table IV (the standard normal distribution table), the z value for .01 area in the left tail is -2.33. Note that the z value is negative because it is a left-tailed test. Thus, we will reject H_0 if the observed value of z is -2.33 or lower, and we will not reject H_0 if the observed value of z is greater than -2.33. Figure 15.4 shows the rejection and nonrejection regions.

Figure 15.4



Step 4. *Calculate the value of the test statistic.*

In our example, 51 phone bills out of 90 are below the hypothesized median, 38 are above it, and 1 is exactly equal to the median. When we perform this test, we drop the value or values that are exactly equal to the median. Thus, after dropping one value that is equal to the median, our sample size is $51 + 38 = 89$; that is, $n = 89$. Let a phone bill below the median be represented by a minus sign and one above the median by a plus sign. Then, in these 89 bills there are 51 minus signs (for values less than the median) and 38 plus signs (for values greater than the median). If the given claim is true, we would expect (about) half plus signs and half minus signs. Let p be the proportion of plus signs in 89. Then, we would expect $p = .50$ if H_0 is true. Hence, the mean and the standard deviation of the binomial distribution are calculated as follows:

$$n = 89 \quad p = .50 \quad q = 1 - p = .50$$
$$\mu = np = 89(.50) = 44.50$$
$$\sigma = \sqrt{npq} = \sqrt{89(.50)(.50)} = 4.71699057$$

In our example, 51 phone bills are below the median and 38 are above the median. Because it is a left-tailed test, $X = 38$, which is the smaller of the two numbers (51 and 38). Consequently, the z value is calculated as follows. Note that we have added .5 to X because the value of X is less than $n/2$, which is $89/2 = 44.5$.

$$z = \frac{(X + .5) - \mu}{\sigma} = \frac{(38 + .5) - 44.50}{4.71699057} = -1.27$$

Step 5. *Make a decision.*

Because $z = -1.27$ is greater than the critical value of $z = -2.33$, we do not reject H_0. Hence, we conclude that the company's claim that the median phone bill is at least $70 seems to be true. ■

▶ Observation

Note that in Example 15–4 there were 51 minus signs and 38 plus signs. We assigned the smaller of these two numbers to X so that $X = 38$ to calculate the observed value of z. We did so to obtain a negative value of the observed z because the test is left-tailed and the critical value of z is negative. If we assigned 51 as the value of X, we would obtain $z = +1.27$ as the observed value of z, which does not make sense. Let X_1 be the number of plus signs and X_2 the number of minus signs in a test about the median. Then, we can establish the following rules to calculate the observed value of X.

1. If the test is two-tailed, it does not matter which of the two values, X_1 and X_2, is assigned to X to calculate the observed value of z.
2. If the test is left-tailed, X should be assigned a value equal to the smaller of the values of X_1 and X_2.
3. If the test is right-tailed, X should be assigned a value equal to the larger of the values of X_1 and X_2.

Note that the rule to calculate the observed value of z here is the same as explained on page 635 for the large-sample case for a test of hypothesis about the preference for categorical data.

15.1.3 Tests About the Median Difference Between Paired Data

We can use the sign test to perform a test of hypothesis about the difference between the medians of two dependent populations using the data obtained from paired samples. We learned in Section 10.4 of Chapter 10 that two samples are paired samples when, for each data value collected from one sample, there is a corresponding data value collected from the second sample, and both data values are collected from the same source. In this section we discuss the small-sample and the large-sample cases to conduct such tests.

The Small-Sample Case

If $n \leq 25$, we use the binomial probability distribution to perform a test about the difference between the medians of paired data. In such a case, Table VIII is used to find the critical values of the test statistic. Example 15–5 illustrates this procedure.

■ EXAMPLE 15–5

Performing sign test about the median of paired differences: small samples.

A researcher wanted to find the effects of a special diet on systolic blood pressure in adults. She selected a sample of 12 adults and put them on this dietary plan for three months. The following table gives the systolic blood pressure of each adult before and after the completion of the plan.

Before	210	185	215	198	187	225	234	217	212	191	226	238
After	196	192	204	193	181	233	208	211	190	186	218	236

Using the 2.5% significance level, can we conclude that the dietary plan reduces the median systolic blood pressure of adults?

Solution We find the sign of the difference between the two blood pressure readings of each adult by subtracting the blood pressure after completion of the dietary plan from the blood pressure before the plan. A plus sign indicates that the plan reduced that person's blood pressure and a minus sign means that it increased blood pressure. Table 15.3 gives the signs of the differences.

Table 15.3

Before	210	185	215	198	187	225	234	217	212	191	226	238
After	196	192	204	193	181	233	208	211	190	186	218	236
Sign of difference (before − after)	+	−	+	+	+	−	+	+	+	+	+	+

Next we perform the five steps for testing the hypothesis.

Step 1. *State the null and alternative hypotheses.*

Let M denote the difference in median blood pressure readings before and after the dietary plan. The null and alternative hypotheses are as follows:

H_0: $M = 0$ (The dietary plan does not reduce the median blood pressure)

H_1: $M > 0$ (The dietary plan reduces the median blood pressure)

The alternative hypothesis is that the dietary plan reduces the median blood pressure, which means that the median systolic blood pressure of all adults after the completion of the dietary plan is lower than the median systolic blood pressure before the completion of the dietary plan. In this case, the median of the paired differences will be greater than zero.

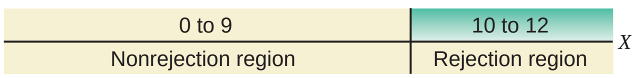
Step 2. *Select the distribution to use.*

The sample size is small (that is, $n = 12 < 25$), and we do not know the shape of the distribution of the population of paired differences. Hence, we use the sign test with the binomial probability distribution.

Step 3. *Determine the rejection and nonrejection regions.*

Because the test is right-tailed, $n = 12$, and $\alpha = .025$, the critical value of X from Table VIII is 10. Note that we use the upper critical value of X in Table VIII because, as indicated by the sign in H_1, the test is right-tailed. Thus, we will reject the null hypothesis if the observed value of X is greater than or equal to 10, and we will not reject H_0 otherwise. The rejection and nonrejection regions are shown in Figure 15.5.

Figure 15.5



Step 4. *Calculate the value of the test statistic.*

In our sample data, the blood pressure of 10 adults decreases and that of 2 adults increases after the dietary plan. Note that there are 10 plus signs and 2 minus signs in Table 15.3. Whenever the test is right-tailed, the observed value of X is equal to the larger of these two numbers. Thus, for our example, observed value of $X = 10$.

Step 5. *Make a decision.*

Because the observed value of $X = 10$ falls in the rejection region, we reject H_0. Hence, we conclude that the dietary plan reduces the median blood pressure of adults. ■

The Large-Sample Case

In Example 15–5, we used Table VIII to find the critical value of the test statistic X. However, Table VIII goes up to $n = 25$ only. If $n > 25$, we can use the normal distribution as an approximation to the binomial distribution to conduct a test about the difference between the medians of paired data. The following example illustrates such a case.

■ EXAMPLE 15–6

Performing sign test about the median of paired differences: large samples.

Many students suffer from math anxiety. A statistics professor offered a two-hour lecture on math anxiety and ways to overcome it. A total of 42 students attended this lecture. The students were given similar statistics tests before and after the lecture. Thirty-three of the 42 students scored higher on the test after the lecture, 7 scored lower after the lecture, and 2 scored the same on both tests. Using the 1% significance level, can you conclude that the median score of students increases as a result of attending this lecture? Assume that these 42 students constitute a random sample of all students who suffer from math anxiety.

Solution Let M be the median of the paired differences between scores of students before and after the test, where a paired difference is obtained by subtracting the score after the lecture from the score before the lecture. In other words,

$$\text{Paired difference} = \text{Score before} - \text{Score after}$$

Thus, a positive paired difference means that the score before the lecture is higher than the score after the lecture for that student, and a negative paired difference indicates that the score before the lecture is lower than the score after the lecture for that student. Thus, there are 33 minus signs, 7 plus signs, and 2 zeros.

Step 1. *State the null and alternative hypotheses.*

$$H_0\colon M = 0 \quad \text{(The lecture does not increase the median score)}$$

$$H_1\colon M < 0 \quad \text{(The lecture increases the median score)}$$

The alternative hypothesis is that the lecture increases the median score, which means that the median score after the lecture is higher than the median score before the lecture. In this case, the median of the paired differences will be less than zero.

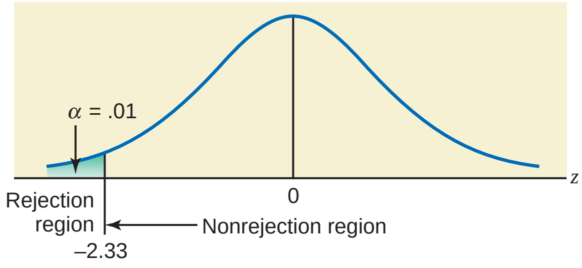
Step 2. *Select the distribution to use.*

Here, $n = 40$. Note that to find the sample size, we drop the students whose score did not change. Because $n > 25$, we can use the normal distribution to test this hypothesis about the median of paired differences.

Step 3. *Determine the rejection and nonrejection regions.*

The test is left-tailed and $\alpha = .01$. From Table IV, the critical value of z for .01 area in the left tail is -2.33. Thus, the decision rule is that we will reject the null hypothesis if the observed value of z is -2.33 or smaller, and we will not reject H_0 otherwise. The rejection and nonrejection regions are shown in Figure 15.6.

Figure 15.6



Step 4. *Calculate the value of the test statistic.*

If the null hypothesis is true (that is, the lecture does not increase the median score), then we would expect (about) half of the students to score higher and the other half to score lower after the lecture than before. Thus, we would expect (about) half plus signs and half minus signs in the population. In other words, if p is the proportion of plus signs, we would expect $p = .50$ when H_0 is true. Hence, the mean and standard deviation of the binomial distribution are

$$\mu = np = 40(.50) = 20$$

$$\sigma = \sqrt{npq} = \sqrt{40(.50)(.50)} = 3.16227766$$

In our example, 33 of the students scored higher after the lecture and 7 scored lower after the lecture. Thus, there are 33 minus signs and 7 plus signs. We assign the smaller of these two values to X when the test is left-tailed. Here $X = 7$, and the observed value of z is calculated as follows. Here, because the value of X is less than $n/2$, we add .5 to X.

$$z = \frac{(X + .5) - \mu}{\sigma} = \frac{(7 + .5) - 20}{3.16227766} = -3.95$$

Step 5. *Make a decision.*

Because the observed value of $z = -3.95$ is less than the critical value of $z = -2.33$, it falls in the rejection region. Consequently, we reject H_0 and conclude that attending the math anxiety lecture increases the median test score. ■

Remember ▶ Again, remember that if the test is left-tailed, X is assigned the value equal to the smaller number of plus or minus signs. On the other hand, if the test is right-tailed, X is assigned the value equal to the larger number of plus or minus signs. Note that the rule to calculate the observed value of z here is the same as explained on page 635 for the large-sample case for a test of hypothesis about the preference for categorical data.

EXERCISES

■ CONCEPTS AND PROCEDURES

15.1 Briefly explain the meaning of categorical data and give two examples.

15.2 When we use the sign test for categorical data, how large a sample size is required to permit the use of the normal distribution for determining the rejection region?

15.3 When we use the sign test for the median of a single population, how small must the sample size be to require the use of Table VIII ?

15.4 When we use the sign test for the difference between the medians of two dependent populations, how large must n be for the large-sample case?

15.5 Determine the rejection region for each of the following sign tests for categorical data.

a. H_0: $p = .50$, H_1: $p > .50$, $n = 15$, $\alpha = .05$
b. H_0: $p = .50$, H_1: $p \neq .50$, $n = 20$, $\alpha = .01$
c. H_0: $p = .50$, H_1: $p < .50$, $n = 30$, $\alpha = .05$

15.6 In each case below, n is the sample size, p is the proportion of the population possessing a certain characteristic, and X is the number of items in the sample that possess that characteristic. In each case, perform the appropriate sign test using $\alpha = .05$.

a. $n = 14$, $X = 10$, H_0: $p = .50$, H_1: $p > .50$
b. $n = 10$, $X = 1$, H_0: $p = .50$, H_1: $p \neq .50$
c. $n = 30$, $X = 12$, H_0: $p = .50$, H_1: $p < .50$
d. $n = 27$, $X = 20$, H_0: $p = .50$, H_1: $p > .50$

15.7 In each case below, n is the sample size and X is the appropriate number of plus or minus signs as defined in Section 15.1.2. In each case, perform the appropriate sign test using $\alpha = .05$.

a. $n = 10$, $X = 8$, H_0: Median $= 28$, H_1: Median > 28
b. $n = 11$, $X = 1$, H_0: Median $= 100$, H_1: Median < 100
c. $n = 26$, $X = 3$, H_0: Median $= 180$, H_1: Median $\neq 180$
d. $n = 30$, $X = 6$, H_0: Median $= 55$, H_1: Median < 55

15.8 In each case below, M is the difference between two population medians, n is the sample size, and X is the appropriate number of plus or minus signs as defined at the end of Section 15.1.3. In each case, perform the appropriate sign test using $\alpha = .01$.

a. $n = 20$, $X = 6$, H_0: $M = 0$, H_1: $M < 0$
b. $n = 8$, $X = 8$, H_0: $M = 0$, H_1: $M > 0$
c. $n = 29$, $X = 4$, H_0: $M = 0$, H_1: $M \neq 0$

APPLICATIONS

15.9 In Pine Grove, the city water is safe to drink but some people think it has a slightly unpleasant taste due to chemical treatment. Some residents prefer to buy bottled water (B) but others drink the city water (C). A random sample of 12 residents is taken. Their preferences are shown here.

B C B C C B C C C C B C

At the 5% significance level, can you conclude that the residents of Pine Grove prefer either type of drinking water over the other?

15.10 A consumer organization wanted to compare two rival brands of infant car seats, Brand A and Brand B. Fifteen families, each with a child under 12 months of age, were selected at random. Each family tested each of the two brands of car seats for one week. The order in which each family tried the two brands was decided by a coin toss. At the end of two weeks, each family indicated which brand it preferred. Their preferences are listed here. The 0 indicates that one family had no preference.

A A A B A A B A A A A 0 A B A

At the 5% level of significance, can you conclude that families prefer Brand A over Brand B?

15.11 Twenty randomly chosen loyal drinkers of JW's Beer are tested to see if they can distinguish between JW's and its chief rival. Each of the 20 drinkers is given two unmarked cups, one containing JW's and the other containing the rival brand. Thirteen of the drinkers correctly indicate which cup contains JW's, but the other seven are incorrect. At the 2.5% level of significance, can you conclude that drinkers of JW's Beer are more likely to correctly identify it than not?

15.12 Three weeks before an election for state senator, a poll of 200 randomly selected voters shows that 95 voters favor the Republican candidate, 85 favor the Democratic candidate, and the remaining 20 have no opinion. Using the sign test, can you conclude that voters prefer one candidate over the other? Use $\alpha = .01$.

15.13 One hundred randomly chosen adult residents of North Dakota are asked whether they would prefer to live in another state or to stay in North Dakota. Of these 100 adults, 55 indicate that they would like to move to another state, 41 would prefer to stay, and 4 have no preference. At the 2.5% level of significance, can you conclude that less than half of all adult residents of North Dakota would prefer to stay?

15.14 Three hundred randomly chosen doctors were asked, Which is the most important single factor in weight control: diet or exercise? Of these 300 doctors, 162 felt that diet was more important, 117 favored exercise, and 21 thought that diet and exercise were equally important. At the 1% level of significance, can you conclude that for all doctors, the number who favor diet exceeds the number who favor exercise?

15.15 In a Gallup Poll of adults taken December 6–9, 2001, 42% reported that they frequently experienced stress in their daily lives (*USA TODAY*, January 24, 2002). Suppose that in a recent sample of 700 adults, 370 indicated that they frequently experience such stress. Using the sign test with $\alpha = .01$, can you conclude that currently more than half of all adults frequently experience stress in their daily lives?

15.16 A past study claims that adults in the United States spend a median of 18 hours a week on leisure activities. A researcher took a sample of 10 adults and asked them how many hours they spend per week on leisure activities. She obtained the following data:

14	25	22	38	16	26	19	23	41	33

Using $\alpha = .05$, can you conclude that the median amount of time spent per week on leisure activities by all adults is more than 18 hours?

15.17 The manager of a soft-drink bottling plant wants to see if the median amount of soda in 12-ounce bottles differs from 12 ounces. Ten filled bottles are selected at random from the bottling machine, and the amount of soda in each is carefully measured. The results (in ounces) follow:

12.10	11.95	12.00	12.01	12.02	12.05	12.02	12.03	12.04	12.06

Using the 5% level of significance, can you conclude that the median amount of soda in all such bottles differs from 12 ounces?

15.18 According to the annual *USA TODAY*/NFL salary survey, the median salary of offensive linemen in the National Football League (NFL) was $589,133 in 2001 (*USA TODAY*, July 29, 2002). Suppose that a recent random sample of 10 NFL offensive linemen yielded the following salaries (in thousands of dollars).

700	615	710	805	630	575	900	730	710	695

Using the 5% level of significance, can you conclude that the current median salary of all offensive linemen in the NFL exceeds $589,133?

15.19 A city police department claims that its median response time to 911 calls in the inner city is four minutes or less. Shown below is a random sample of 28 response times (in minutes) to 911 calls in the inner city.

6	5	7	12	2	1.5	3.5	4	10	11	4.5	6	5	8.5
7	15	9	8	3	10	8	4.5	9	4	6	3	6	7.5

Using $\alpha = .01$, can you conclude that the median response time to all 911 calls in the inner city is longer than four minutes?

15.20 According to the American Community Survey conducted during the 2000 census, New Jersey's median household income of $54,226 was the highest among the 50 states (*USA TODAY*, August 6, 2001). Suppose that in a recent random sample of 400 New Jersey households, 220 had incomes higher than $54,226 and 180 had incomes lower than $54,226. Using the sign test at the 2% level of significance, can you conclude that the current median household income in New Jersey differs from $54,226?

15.21 The following numbers are the times served (in months) by 35 prison inmates who were released recently.

37	6	20	5	25	30	24	10	12	20
24	8	26	15	13	22	72	80	96	33
84	86	70	40	92	36	28	90	36	32
72	45	38	18	9					

Using $\alpha = .01$, test the null hypothesis that the median time served by all such prisoners is 42 months against the alternative hypothesis that the median time served is less than 42 months.

15.22 Twelve sixth-grade boys who are underweight are put on a special diet for one month. Each boy is weighed before and after the one-month dietary regime. The weights (in pounds) of these boys are recorded here.

Before	65	63	71	60	66	72	78	74	58	59	77	65
After	70	68	75	60	69	70	81	81	66	56	79	71

Can you conclude that this diet increases the median weight of all such boys? Use the 2.5% level of significance. Assume that these 12 boys constitute a random sample of all underweight sixth-grade boys.

15.23 Refer to Exercise 10.52 of Chapter 10. The following table shows the self-confidence test scores of seven employees before and after they attended a course on building self-confidence.

Before	8	5	4	9	6	9	5
After	10	8	5	11	6	7	9

At the 5% level of significance, can you conclude that attending this course increases the median self-confidence test score of all employees?

15.24 The manager at a large factory suspects that the night-shift workers use more hours of sick leave than the day-shift workers. The workers at this factory are rotated between shifts. Each worker works the day shift for two months, then works the night shift for two months, then goes back to the day shift for two months, and so on. The manager at the factory selected 12 workers randomly and recorded the total number of hours of sick leave each of these workers used during the two months of day shift and then during the two months of night shift. The results are given in the following table.

Day shift	20	32	12	24	16	0	22	8	10	38	16	12
Night shift	16	56	0	28	36	24	40	29	30	26	32	20

Using the 5% level of significance, can you conclude that the median number of hours of sick leave used by workers is lower for the day shift than for the night shift?

15.25 At a large bicycle factory, employees are paid by the hour to assemble bicycles. The plant manager decides to test a modified-piecework payment schedule, whereby each worker will be paid a lower hourly wage plus an additional amount for each bicycle assembled. The manager randomly selects 27 workers and places them on the new payment schedule. For each worker in the sample, the number of bicycles assembled during the last week under the old payment system is recorded, and then the number of bicycles assembled during the first week under the new system is recorded. Nineteen of the workers assembled more bicycles under the new system, seven assembled fewer, and one assembled the same number. Using the 2% level of significance, can you conclude that the median number of bicycles assembled by all such workers is the same under both payment systems?

15.26 A researcher suspects that two medical laboratories, A and B, tend to give different results when determining the cholesterol content of blood samples. The researcher obtains blood samples from 30 randomly selected adults and divides each sample into two parts. One part of each blood sample is sent to Lab A, the other part to Lab B. Each lab determines the cholesterol content of each of its 30 samples and reports the results to the researcher. The following table gives the cholesterol levels (in milligrams per hundred milliliters) reported by the two labs.

Sample	Lab A	Lab B	Sample	Lab A	Lab B
1	135	137	16	214	202
2	202	195	17	255	242
3	239	250	18	233	217
4	210	202	19	246	231
5	180	185	20	292	262
6	195	195	21	229	212
7	188	177	22	170	172
8	200	204	23	261	243
9	320	300	24	310	281
10	290	269	25	302	277
11	285	271	26	283	264
12	210	216	27	221	199
13	185	176	28	208	211
14	194	184	29	344	321
15	181	182	30	170	164

Can you conclude that the median cholesterol level for all such adults as determined by Lab A is higher than that determined by Lab B? Use a significance level of 1%.

15.27 A dairy agency wants to test a hormone that may increase cows' milk production. Some members of the group fear that the hormone could actually decrease production, so a "matched pairs" test is arranged. Thirty randomly selected cows are given the hormone, and their milk production is recorded for four weeks. Each of these 30 cows is matched with another cow of similar size, age, and prior record of milk production. This second group of 30 cows do not receive the hormone. The milk production of these cows is recorded for the same time period. In 19 of these 30 pairs, the cow taking the hormone produced more milk; in 9 of the pairs, the cow taking the hormone produced less; and in 2 of the pairs, there was no difference. Using the 5% level of significance, can you conclude that the hormone changes the median milk production of such cows?

15.2 The Wilcoxon Signed-Rank Test for Two Dependent Samples

The **Wilcoxon signed-rank test** for two dependent (paired) samples is used to test whether or not the two populations from which these samples are drawn are identical. We can also test the alternative hypothesis that one population distribution lies to the left or to the right of the other. Actually, the null hypothesis in this test states that the medians of the two population distributions are equal. The alternative hypothesis states that the medians of the two populations are not equal, or that the median of the first population is less than that of the second population, or that the median of the first population is greater than that of the second population. This test is an alternative to the paired-samples test discussed in Section 10.4.2 of Chapter 10. In that section, we assumed that the paired differences have a normal distribution. Here, in the Wilcoxon signed-rank test, we do not make that assumption. In this test, we rank the absolute differences between the pairs of data values collected from two samples and then assign them the sign based on which of the paired data values is larger. Then we compare the sums of the ranks with plus and minus signs and make a decision.

The Small-Sample Case

If the *sample size is 15 or smaller*, we find the critical value of the test statistic, denoted by T, from Table IX (given at the end of this chapter) which gives the critical values of T for the Wilcoxon signed-rank test. We also calculate the observed value of the test statistic differently in this test. However, when $n > 15$, we can use the normal distribution to perform the test. Example 15–7 describes the small-sample case for the Wilcoxon signed-rank test.

■ EXAMPLE 15–7

Performing the Wilcoxon signed-rank test for two dependent populations: small samples.

A private agency claims that the crash course it offers significantly increases the writing speed of secretaries. The following table gives the writing speeds of eight secretaries before and after they attended this course.

Before	84	75	88	91	65	71	90	75
After	97	72	93	110	78	69	115	75

Using the 2.5% significance level, can you conclude that attending this course increases the writing speed of secretaries? Use the Wilcoxon signed-rank test.

Solution We use the five steps to make the hypothesis test.

Step 1. *State the null and alternative hypotheses.*

H_0: The crash course does not increase the writing speed of secretaries

H_1: The crash course does increase the writing speed of secretaries

Note that the alternative hypothesis states that the population distribution of writing speeds of secretaries moves to the right after they attend the crash course. In other words, the center of the population distribution of writing speeds after the crash course is greater than the center of the population distribution of writing speeds before the crash course. If we measure the centers of the two populations by their respective medians, with M_A the median of the population distribution after the course and M_B the median of the population distribution before the course, we can rewrite the two hypotheses as follows:

$$H_0: M_A = M_B$$

$$H_1: M_A > M_B$$

Step 2. *Select the distribution to use.*

We are making a test for paired samples, and the distribution of paired differences is unknown. Since $n < 15$, we use the Wilcoxon signed-rank test procedure for the small-sample case.

Step 3. *Determine the rejection and nonrejection regions.*

As mentioned earlier, we denote the test statistic in this case by T. The critical value of T is found from Table IX, which lists the critical values of T for the Wilcoxon signed-rank test for small samples ($n \le 15$). Our test is right-tailed because the alternative hypothesis is that the "after" distribution lies to the right of the "before" distribution. Also, $\alpha = .025$ and $n = 7$. Note that for one pair of data, both values are the same, 75. We drop such cases when determining the sample size for the test. From Table IX, the critical value of T is 2. Thus, our decision rule will be: Reject H_0 if the observed value of T is less than or equal to the critical value of T, which is 2. Note that in the Wilcoxon signed-rank test, the null hypothesis is rejected if the observed value of T is less than or equal to the critical value of T. This rule is true for a two-tailed, a right-tailed, or a left-tailed test. The observed value of T is calculated differently, depending on whether the test is two-tailed or one-tailed. This is explained in the next step. Figure 15.7 shows the rejection and nonrejection regions.

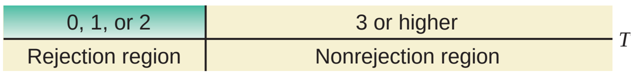


Figure 15.7

Decision Rule For the Wilcoxon signed-rank test for small samples ($n \le 15$), the critical value of T is obtained from Table IX. Note that in the Wilcoxon signed-rank test, the decision rule is to reject the null hypothesis if the observed value of T is less than or equal to the critical value of T. This rule is true for a two-tailed, a right-tailed, or a left-tailed test.

Step 4. *Calculate the value of the test statistic.*

The observed value of the test statistic, T, is calculated as follows. The given data on writing speeds before and after the course are reproduced in the first two columns of Table 15.4.

Table 15.4

Before	After	Differences (Before − After)	Absolute Differences	Ranks of Differences	Signed Ranks
84	97	−13	13	4.5	−4.5
75	72	+3	3	2	+2
88	93	−5	5	3	−3
91	110	−19	19	6	−6
65	78	−13	13	4.5	−4.5
71	69	+2	2	1	+1
90	115	−25	25	7	−7
75	75	0	0	—	—

1. We obtain the differences column by subtracting each data value after the course from the corresponding data value before the course. Thus,

   Difference = Writing speed before the course − Writing speed after the course

   These differences are listed in the third column of Table 15.4.

2. In the fourth column, we write the absolute values of differences. In other words, the numbers in the fourth column of the table are the same as those in the third column, but without plus and minus signs.

3. Next, we rank the absolute differences listed in the fourth column from lowest to highest. These ranks are listed in the fifth column. Note that the difference of zero is not ranked and is dropped from the sample. Among the remaining absolute differences, the smallest difference is 2, which is assigned a rank of 1. The next smallest absolute difference is 3, which is assigned a rank of 2. Next, the absolute difference of 5 is given a rank of 3. Then, two absolute differences have the same value, which is 13. We assign the average of the next two ranks, $(4 + 5)/2 = 4.5$, to these two values. Thus, as a rule, whenever some of the absolute differences have the same value, they are all assigned the average of their ranks.

4. In the last column of Table 15.4, we write the ranks of the fifth column with the signs of the corresponding paired differences. For example, the first difference of -13 has a minus sign in the third column. Consequently, we assign a minus sign to its rank of 4.5 in the sixth column. The second difference of 3 has a positive sign. Hence, its rank of 2 is assigned a positive sign.

5. Next, we add all the positive ranks and we add the absolute values of the negative ranks separately. Thus, we obtain:

   Sum of the positive ranks $= 2 + 1 = 3$

   Sum of the absolute values of the negative ranks $= 4.5 + 3 + 6 + 4.5 + 7 = 25$

The observed value of the test statistic is determined as shown in the next box.

Observed Value of the Test Statistic T

I. If the test is two-tailed with the alternative hypothesis that the two distributions are not the same, then the observed value of T is given by the smaller of the two sums, the sum of the positive ranks and the sum of the absolute values of the negative ranks. We will reject H_0 if the observed value of T is less than or equal to the critical value of T.

II. If the test is right-tailed with the alternative hypothesis that the distribution of *after* values is to the right of the distribution of *before* values, then the observed value of T is given by the sum of the values of the positive ranks. We will reject H_0 if the observed value of T is less than or equal to the critical value of T.

III. If the test is left-tailed with the alternative hypothesis that the distribution of *after* values is to the left of the distribution of *before* values, then the observed value of T is given by the sum of the absolute values of the negative ranks. We will reject H_0 if the observed value of T is less than or equal to the critical value of T.

Remember, for the above to be true, the paired difference is defined as the *before* value minus the *after* value. In other words, the differences are obtained by subtracting the *after* values from the *before* values.

Our example is a right-tailed test. Hence,

$$\text{Observed value of } T = \text{sum of the positive ranks} = 3$$

Step 5. *Make a decision.*

Whether the test is two-tailed, left-tailed, or right-tailed, we will reject the null hypothesis if:

$$\text{Observed value of } T \leq \text{Critical value of } T$$

where the observed value of T is calculated as explained in Step 4. In this example, the observed value of T is 3 and the critical value of T is 2. Because the observed value of T is greater than the critical value of T, we do not reject H_0. Hence, we conclude that the crash course does not seem to increase the writing speed of secretaries.

The Large-Sample Case

If $n > 15$, we can use the normal distribution to make a test of hypothesis about the paired differences. Example 15–8 illustrates the procedure for making such a test.

■ EXAMPLE 15–8

Performing the Wilcoxon signed-rank test for paired populations: large samples.

The manufacturer of a gasoline additive claims that the use of its additive increases gasoline mileage. A random sample of 25 cars was selected, and these cars were driven for one week without the gasoline additive and then for one week with the additive. Then, the miles per gallon (mpg) were estimated for these cars without and with the additive. Next, the paired differences were calculated for these 25 cars, where a paired difference is defined as

$$\text{Paired difference} = \text{mpg without additive} - \text{mpg with additive}$$

The differences were positive for 4 cars, negative for 19 cars, and zero for 2 cars. First, the absolute values of the paired differences were ranked, and then these ranks were assigned the signs of the corresponding paired differences. The sum of the ranks of the positive paired differences was 58, and the sum of the absolute values of the ranks of the negative paired differences was 218. Can you conclude that the use of the additive increases gasoline mileage? Use the 1% significance level.

Solution We perform the five steps to conduct this test of hypothesis.

Step 1. *State the null and alternative hypotheses.*

We are to test whether or not the gasoline additive increases gasoline mileage. This will be true if the distribution of gasoline mileages with the additive lies to the right of the distribution of gasoline mileage without the additive. The median mileage with the additive will be higher than the median mileage without the additive. Let M_A and M_B be the median mileage after (with) and before (without) the gasoline additive. Then, the null and the alternative hypotheses can be written as follows:

$$H_0\colon M_A = M_B$$
$$H_1\colon M_A > M_B$$

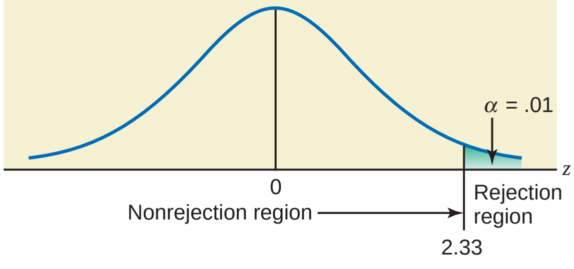
Step 2. *Select the distribution to use.*

We are given information about the sums of positive and negative ranks. The sample size is greater than 15. We use the Wilcoxon signed-rank test procedure with the normal distribution approximation.

Step 3. *Determine the rejection and nonrejection regions.*

We are using the normal distribution as an approximation to make this test. Hence, we will find the critical value of z from Table IV in Appendix C. The test is right-tailed. The significance level is .01, which gives the area to the left of critical point as $1 - .01 = .9900$. Therefore, the critical value of z is 2.33. The rejection and nonrejection regions are shown in Figure 15.8.

Figure 15.8



Step 4. *Calculate the value of the test statistic.*

Because the sample size is larger than 15, the test statistic T follows (an approximate) normal distribution.

Observed Value of z In a Wilcoxon signed-rank test for two dependent samples, when the sample size is large ($n > 15$), the observed value of z for the test statistic T is calculated as

$$z = \frac{T - \mu_T}{\sigma_T}$$

where
$$\mu_T = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

The value of T that is used to calculate the value of z is determined based on the alternative hypothesis, as explained next.

1. If the test is two-tailed with the alternative hypothesis that the two distributions are not the same, then the value of T may be equal to either of the two sums, the sum of the positive ranks or the sum of the absolute values of the negative ranks. We will reject H_0 if the observed value of z falls in either of the rejection regions.
2. If the test is right-tailed with the alternative hypothesis that the distribution of *after* values is to the right of the distribution of *before* values, then the value of T is equal to the sum of the absolute values of the negative ranks. We will reject H_0 if the observed value of z is greater than or equal to the critical value of z.
3. If the test is left-tailed with the alternative hypothesis that the distribution of *after* values is to the left of the distribution of *before* values, then the value of T is equal to the sum of the absolute values of the negative ranks. We will reject H_0 if the observed value of z is less than or equal to the critical value of z.

Remember, for the above to be true, the paired difference is defined as the *before* value minus the *after* value. In other words, the differences are obtained by subtracting the *after* values from the *before* values. Also, note that whether the test is right-tailed or left-tailed, the value of T in both cases is equal to the sum of the absolute values of the negative ranks.

Using the given information, we calculate the values of μ_T and σ_T and the observed value of z as follows. Note that here $n = 23$ because two of the paired differences are zero.

$$\mu_T = \frac{n(n+1)}{4} = \frac{23(23+1)}{4} = 138$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{23(23+1)(46+1)}{24}} = 32.87856445$$

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{218 - 138}{32.87856445} = 2.43$$

Step 5. *Make a decision.*

The observed value of $z = 2.43$ falls in the rejection region. Hence, we reject the null hypothesis and conclude that the gasoline additive increases mileage. ■

EXERCISES

CONCEPTS AND PROCEDURES

15.28 When would you use the Wilcoxon signed-rank test procedure instead of the paired-samples test of Chapter 10?

15.29 Explain how the null hypothesis is usually stated in the Wilcoxon signed-rank test.

15.30 How are ranks assigned to two or more absolute differences that have the same value in the Wilcoxon signed-rank test?

15.31 Determine the rejection region for the Wilcoxon signed-rank test for each of the following. Indicate whether the rejection region is based on T or z.

a. $n = 10$, H_0: $M_A = M_B$, H_1: $M_A > M_B$, $\alpha = .05$
b. $n = 12$, H_0: $M_A = M_B$, H_1: $M_A \neq M_B$, $\alpha = .01$
c. $n = 20$, H_0: $M_A = M_B$, H_1: $M_A < M_B$, $\alpha = .025$
d. $n = 30$, H_0: $M_A = M_B$, H_1: $M_A > M_B$, $\alpha = .01$

15.32 In each case, perform the Wilcoxon signed-rank test.

a. $n = 8$, $T = 5$, left-tailed test using $\alpha = .05$
b. $n = 15$, $T = 20$, right-tailed test using $\alpha = .01$
c. $n = 25$, $T = 51$, two-tailed test using $\alpha = .02$
d. $n = 36$, $T = 238$, left-tailed test using $\alpha = .01$

APPLICATIONS

15.33 Refer to Exercise 10.96 of Chapter 10, which deals with Gamma Corporation's installation of governors on its salespersons' cars to regulate their speeds. The following table gives the number of contacts made by each of seven randomly selected sales representatives during the week before governors were installed and the number of contacts made during the week after installation.

Salesperson	A	B	C	D	E	F	G
Before	50	63	42	55	44	65	66
After	49	60	47	51	50	60	58

a. Using the Wilcoxon signed-rank test at the 5% level of significance, can you conclude that the use of governors tends to reduce the number of contacts made per week by Gamma Corporation's sales representatives?
b. Compare your conclusions of part a with the result of the hypothesis test that was performed (using the t distribution) in Exercise 10.96.

15.34 Refer to Exercise 10.96 of Chapter 10. The following table gives the gas mileage (in miles per gallon) for each of seven randomly selected sales representatives' cars during the week before governors were installed and the gas mileage in the week after installation.

Salesperson	A	B	C	D	E	F	G
Before	25	21	27	23	19	18	20
After	26	24	26	25	24	22	23

a. Using the Wilcoxon signed-rank test at the 5% level of significance, can you conclude that the use of governors tends to increase the median gas mileage for Gamma Corporation's sales representatives?
b. Compare your conclusion of part a with the result of the hypothesis test that was performed (using the t distribution) in Exercise 10.96.

15.35 Refer to Exercise 15.23. The following table shows the self-confidence test scores of seven employees before and after they attended a course designed to build self-confidence.

Before	8	5	4	9	6	9	5
After	10	8	5	11	6	7	9

a. Using the Wilcoxon signed-rank test at the 5% level of significance, can you conclude that attending this course increases the median self-confidence test score of employees?
b. Compare your conclusion of part a with the result of Exercise 15.23.

15.36 Refer to Exercise 15.25, which compares productivity of 27 bicycle assemblers under an hourly payment system and under a modified-piecework payment scheme. The paired difference for each assembler was calculated by subtracting the number of bicycles assembled during the first week under the new payment system from the number of bicycles assembled during the last week under the hourly wage system. These paired differences are positive for 7 assemblers, negative for 19, and zero for 1 assembler. The sum of the ranks of the positive paired differences is 61, and the sum of the absolute values of the ranks of the negative paired differences is 290.

a. Using the Wilcoxon signed-rank test at the 2% level of significance, can you conclude that the median number of bicycles assembled by all such assemblers is the same under both payment systems?

b. Compare your conclusion of part a with the result of the sign test performed in Exercise 15.25.

15.37 Twenty randomly selected adults who describe themselves as "couch potatoes" were given a six-week course in physical fitness. Before starting the course, each adult took a two-mile hike on the same trail. The time required to complete the hike was recorded for each adult. After finishing the course, they all took the same hike again, and their times were recorded again. The following table lists the times recorded (in minutes) before and after the course for each of the 20 adults.

Before	After	Before	After	Before	After
41	37	64	55	100	78
91	71	37	31	48	40
35	30	54	57	50	48
58	64.5	70	59	94	102
45	44	40	33	42.5	40
48.5	44	78	70.5	75	63
84	78	66	56		

Does the fitness course appear to reduce the time required to complete the two-mile hike? Use the Wilcoxon signed-rank test at the 2.5% level of significance.

15.38 Many adults in the United States have accumulated excessive balances on their credit cards. Ninety such adults were randomly chosen to participate in a group therapy program designed to reduce their debts. Each adult's total credit card balance was recorded twice: before the program began and three months after the program ended. The paired difference was then calculated for each adult by subtracting the balance after the program ended from the balance before the program began. These paired differences were positive for 49 adults and negative for 41 adults. The sum of the ranks of the positive paired differences was 2507, and the sum of the absolute values of the ranks of the negative paired differences was 1588. Can you conclude that this group therapy program reduces credit card debt? Use the 5% level of significance.

15.3 The Wilcoxon Rank Sum Test for Two Independent Samples

In Chapter 10 we discussed tests of hypotheses about the difference between the means of two independent populations using the normal and t distributions. In this section we compare two independent populations using the results obtained from samples drawn from these populations. In a **Wilcoxon rank sum test**, we assume that the two populations have identical shapes but differ only in location, which is measured by the median. Note that identical shapes do not mean that they have to have a normal distribution. To apply this test, we must be able to rank the given data. Note that the Wilcoxon rank sum test is almost identical to the **Mann-Whitney test**.

In the hypothesis tests discussed in this section, the null hypothesis is usually that the two population distributions are identical. The alternative hypothesis can be that the two population distributions are not identical or that one distribution is to the right of the other or that one distribution is to the left of the other. Assuming that the null hypothesis is true and that the two populations are identical, we rank all the (combined) data values of the two samples as if they were drawn from the same population. Any tied data values are assigned the ranks in the same manner as in the preceding section. Then we sum the ranks for the data values of each sample separately. If the two populations are identical, the ranks should be spread randomly (and evenly) between the two samples. In this case, the sums of the ranks for the two samples should be almost equal, given that the sizes of the two samples are almost the same. However, if one of the two samples contains mostly lower ranks and the other contains mostly higher ranks, then the sums of the ranks for the two samples will be quite different. The larger the difference in the sums of the ranks of the two samples, the more convincing is the evidence that the two populations are not identical and that the null hypothesis is not true.

In this section, we discuss the Wilcoxon rank sum test first for small samples and then for large samples.

The Small-Sample Case

If *the sizes of both samples are 10 or less*, we use the Wilcoxon rank sum test for small samples. Example 15–9 illustrates how the test is performed. To make this test, the population that corresponds to the smaller sample is labeled population 1 and the one that corresponds to the larger sample is called population 2. The respective samples are sample 1 and sample 2. If the sizes of the two samples are equal, either of the two populations can be labeled population 1.

■ EXAMPLE 15–9

Performing the Wilcoxon rank sum test for two independent populations: small samples.

A researcher wants to determine whether the distributions of daily crimes in two cities are identical. The following data give the numbers of violent crimes on eight randomly selected days for City A and on nine days for City B.

City A	12	21	16	8	26	13	19	23	
City B	18	25	14	16	23	19	28	20	31

Using the 5% significance level, can you conclude that the distributions of daily crimes in the two cities are different?

Solution We apply the following five steps to perform the hypothesis test.

Step 1. *State the null and alternative hypotheses.*

We are to test if the two populations are identical or different. Hence, the two hypotheses are as follows:

H_0: The population distributions of daily crimes in the two cities are identical

H_1: The population distributions of daily crimes in the two cities are different

Step 2. *Select the distribution to use.*

Let the distribution of daily crimes in City A be called population 1 (note that it corresponds to the smaller sample) and that in City B be called population 2. The respective samples are called sample 1 and sample 2. Because $n_1 < 10$ and $n_2 < 10$, we use the Wilcoxon rank sum test for small samples.

Step 3. *Determine the rejection and nonrejection regions.*

The test statistic in Wilcoxon's rank sum test is denoted by T. The critical value or values of T are obtained from Table X that appears at the end of this chapter. In this table, T_L gives the lower critical value and T_U gives the upper critical value. If the test is two-tailed, we use both T_L and T_U. For a left-tailed test we use T_L only, and for a right-tailed test we use T_U only.

In our example, the test is two-tailed. Also, $\alpha = .05$, $n_1 = 8$, and $n_2 = 9$. Hence, from Table X, the values of T_L and T_U are 51 and 93, respectively. We will reject the null hypothesis if the observed value of T is either less than or equal to T_L or greater than or equal to T_U. The rejection and nonrejection regions are shown in Figure 15.9. Thus, the decision rule is that we will reject H_0 if either the observed value of $T \leq 51$ or the observed value of $T \geq 93$.

Figure 15.9

51 or lower	52 to 92	93 or higher
Rejection region	Nonrejection region	Rejection region

Step 4. *Calculate the value of the test statistic.*

Table 15.5

City A		**City B**	
Crimes	**Rank**	**Crimes**	**Rank**
12	2	18	7
21	11	25	14
16	5.5	14	4
8	1	16	5.5
26	15	23	12.5
13	3	19	8.5
19	8.5	28	16
23	12.5	20	10
		31	17
	Sum = 58.5		Sum = 94.5

To find the observed value of T, first we rank all the data values of both samples as if they belonged to the same population. Then, we find the sum of the ranks for each sample separately. The observed value of the test statistic T is given by the sum of the ranks for the smaller sample. If the sizes of the samples are the same, we can use either of the rank sums as the observed value of T.

In Table 15.5, we rank all the values of both samples and find the sum of the ranks for each sample. Note that 8 is the smallest data value in both samples. Hence, it is assigned a rank of 1. The next smallest value in both samples is 12, which is assigned a rank of 2. The remaining values are assigned ranks in the same way.

Because $n_1 = 8$ and $n_2 = 9$, the sample size for City A is smaller. Hence, the observed value of T is given by the sum of the ranks for city A. Thus,

$$\text{Observed value of } T = 58.5$$

Step 5. *Make a decision.*

Comparing the observed value of T with T_L and T_U (obtained from Table X in Step 3), we see that the observed value of $T = 58.5$ is between $T_L = 51$ and $T_U = 93$. Hence, we do not reject H_0, and we conclude that the two population distributions seem to be identical. ■

Below we describe the Wilcoxon rank sum test procedure for small samples for two-tailed, right-tailed, and left-tailed tests.

Wilcoxon Rank Sum Test for Small Independent Samples

1. *A two-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the two population distributions are different. The critical values of T, T_L, and T_U for this test are obtained from Table X for the given significance level and sample sizes. The observed value of T is given by the sum of the ranks for the smaller sample. The null hypothesis is rejected if $T \leq T_L$ or $T \geq T_U$. Otherwise, the null hypothesis is not rejected.

Note that if the two sample sizes are equal, the observed value of T is given by the sum of the ranks for either sample.

2. *A right-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the distribution of population 1 (the population that corresponds to the smaller sample) lies to the right of the distribution of population 2. The critical value of T is given by T_U in Table X for the given α for a one-tailed test and the given sample sizes. The observed value of T is given by the sum of the ranks for the smaller sample. The null hypothesis is rejected if $T \geq T_U$. Otherwise, the null hypothesis is not rejected.

Note that if the two sample sizes are equal, the observed value of T is given by the sum of the ranks for sample 1.

3. *A left-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the distribution of population 1 (the population that corresponds to the smaller sample) lies to the left of the distribution of population 2. The critical value of T in this case is given by T_L in Table X for the given α for a one-tailed test and the given sample sizes. The observed value of T is given by the sum of the ranks for the smaller sample. The null hypothesis is rejected if $T \leq T_L$. Otherwise the null hypothesis is not rejected.

Note that if the two sample sizes are equal, the observed value of T is given by the sum of the ranks for sample 1.

The Large-Sample Case

If *either* n_1 *or* n_2 *or both* n_1 *and* n_2 *are greater than 10*, we use the normal distribution as an approximation to the Wilcoxon rank sum test for two independent samples.

Observed Value of *z* In the case of a large sample, the observed value of z is calculated as

$$z = \frac{T - \mu_T}{\sigma_T}$$

Here, the sampling distribution of the test statistic T is approximately normal with mean μ_T and standard deviation σ_T. The values of μ_T and σ_T are calculated as

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \text{and} \quad \sigma_T = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Note that in these calculations sample 1 refers to the smaller sample and sample 2 to the larger sample. However, if the two samples are of the same size, we can label either one sample 1. The value of T used in the calculation of z is given by the sum of the ranks for sample 1.

The critical value or values of z are obtained from Table IV in Appendix C for the given significance level. We will reject the null hypothesis if the observed value of z is in the rejection region. Otherwise, we will not reject H_0. Example 15–10 illustrates the procedure for performing such a test.

■ EXAMPLE 15–10

Performing the Wilcoxon rank sum test for two independent populations: large samples.

A researcher wanted to find out whether job-related stress is lower for college and university professors than for physicians. She took random samples of 14 professors and 11 physicians and tested them for job-related stress. The following data give the stress levels for professors and physicians on a scale of 1 to 20, where 1 is the lowest level of stress and 20 is the highest.

Professors	5	9	4	12	6	15	2	8	10	4	6	11	8	3
Physicians	10	18	12	5	13	18	14	9	6	16	11			

Using the 1% significance level, can you conclude that the job-related stress level for professors is lower than that for physicians?

Solution Because the smaller sample should be labeled sample 1, the sample of 11 physicians will be called sample 1 and that of 14 professors will be called sample 2. The respective populations are populations 1 and 2. Thus, $n_1 = 11$ and $n_2 = 14$. We perform the five steps of the hypothesis test.

Step 1. *State the null and alternative hypotheses.*

We are to test whether or not professors have lower job-related stress than physicians. Because physicians are labeled population 1 and professors population 2, professors will have a lower stress level if the distribution of population 1 is to the right of the distribution of population 2. Thus, we can state the two hypotheses as follows.

H_0: The two population distributions are identical

H_1: The distribution of population 1 is to the right of the distribution of population 2

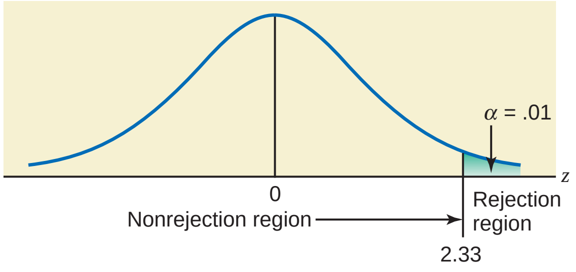
Step 2. *Select the distribution to use.*

Because $n_1 > 10$ and $n_2 > 10$, we use the normal distribution to make this test as the test statistic T follows an approximately normal distribution.

Step 3. *Determine the rejection and nonrejection regions.*

The test is right-tailed and $\alpha = .01$. The area to the left of the critical point under the normal distribution curve is $1 - .01 = .9900$. From Table IV in Appendix C, the critical value of z for .9900 is 2.33. The rejection and nonrejection regions are shown in Figure 15.10. Thus, we will reject H_0 if the observed value of z is 2.33 or greater. Otherwise, we will not reject H_0.

Figure 15.10



Step 4. *Calculate the value of the test statistic.*

Table 15.6 shows the rankings of all the data values for the two samples and the sums of these ranks for each sample separately.

Table 15.6

Physicians		Professors	
Stress Level	**Rank**	**Stress Level**	**Rank**
10	14.5	5	5.5
18	24.5	9	12.5
12	18.5	4	3.5
5	5.5	12	18.5
13	20	6	8
18	24.5	15	22
14	21	2	1
9	12.5	8	10.5
6	8	10	14.5
16	23	4	3.5
11	16.5	6	8
		11	16.5
		8	10.5
		3	2
	Sum = 188.5		Sum = 136.5

Hence, we calculate the value of the test statistic as follows:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{11(11 + 14 + 1)}{2} = 143$$

$$\sigma_T = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{11(14)(11 + 14 + 1)}{12}} = 18.26654501$$

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{188.5 - 143}{18.26654501} = 2.49$$

Thus, the observed value of z is 2.49. Note that in the calculation of z, we used the value of T that belongs to sample 1, which should always be the case.

Step 5. *Make a decision.*

Because the observed value of $z = 2.49$ is greater than the critical value of $z = 2.33$, it falls in the rejection region. Hence, we reject H_0 and conclude that the distribution of population 1 is to the right of the distribution of population 2. Thus, the job-related stress level of physicians is higher than that of professors. This can also be stated as the job-related stress level of professors is lower than that of physicians. ■

Below we describe the Wilcoxon rank sum test procedure for large samples for two-tailed, right-tailed, and left-tailed tests.

Wilcoxon Rank Sum Test for Large Independent Samples When $n_1 > 10$ *or* $n_2 > 10$ (or both samples are greater than 10), the distribution of T (the sum of the ranks of the smaller of the two samples) is approximately normal with mean and standard deviation as follows:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \text{and} \quad \sigma_T = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

For two-tailed, right-tailed, and left-tailed tests, first calculate T, μ_T, σ_T, and the value of the test statistic, $z = (T - \mu_T)/\sigma_T$. If $n_1 = n_2$, T can be calculated from either sample 1 or sample 2.

1. *A two-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the two population distributions are different. At significance level α, the critical values of z are obtained from Table IV in Appendix C. The null hypothesis is rejected if the observed value of z falls in the rejection region.
2. *A right-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the distribution of population 1 (the population with the smaller sample size) lies to the right of the distribution of population 2. At significance level α, the critical value of z is obtained from Table IV in Appendix C. The null hypothesis is rejected if the observed value of z falls in the rejection region.
3. *A left-tailed test*: The null hypothesis is that the two population distributions are identical, and the alternative hypothesis is that the distribution of population 1 (the population with the smaller sample size) lies to the left of the distribution of population 2. At significance level α, the critical value of z is found from Table IV in Appendix C. The null hypothesis is rejected if the observed value of z falls in the rejection region.

EXERCISES

CONCEPTS AND PROCEDURES

15.39 Explain what determines whether to use the Wilcoxon signed-rank test or the Wilcoxon rank sum test.

15.40 Find the rejection region for the Wilcoxon rank sum test in each of the following cases.

a. $n_1 = 7$, $n_2 = 8$, right-tailed test using $\alpha = .05$
b. $n_1 = 10$, $n_2 = 10$, two-tailed test using $\alpha = .10$
c. $n_1 = 18$, $n_2 = 20$, left-tailed test using $\alpha = .05$
d. $n_1 = 25$, $n_2 = 25$, two-tailed test using $\alpha = .01$

15.41 In each of the following cases, perform the Wilcoxon rank sum test.

a. $n_1 = 6$, $n_2 = 7$, $T = 22$, two-tailed test with $\alpha = .05$
b. $n_1 = 10$, $n_2 = 12$, $T = 137$, right-tailed test with $\alpha = .025$
c. $n_1 = 9$, $n_2 = 11$, $T = 68$, left-tailed test with $\alpha = .05$
d. $n_1 = 22$, $n_2 = 23$, $T = 638$, two-tailed test with $\alpha = .01$

APPLICATIONS

15.42 A consumer agency wants to compare the caffeine content of two brands of coffee. Eight jars of each brand are analyzed, and the amount of caffeine found in each jar is recorded as shown in the table.

Brand I	82	77	85	73	84	79	81	82
Brand II	75	80	76	81	72	74	73	78

Using $\alpha = .10$, can you conclude that the two brands have different median caffeine contents per jar?

15.43 In a Winter Olympics trial for women's speed skating, seven skaters use a new type of skate, while eight others use the traditional type. Each skater is timed (in seconds) in the 500-meter event. The results are given in the following table.

New skates	40.5	40.3	39.5	39.7	40.0	39.9	41.5	
Traditional skates	41.0	40.8	40.9	39.8	40.6	40.7	41.1	40.5

Assuming that these 15 skaters make up a random sample of all Olympic-class 500-meter female speed skaters, can you conclude that the new skates tend to produce faster times in this event? Use the 5% level of significance.

15.44 During April–June 2004, the median price of homes sold in Phoenix was $252,400, and the median price in Las Vegas was $255,800. The following table gives the prices (in thousands of dollars) of 9 randomly selected homes in Phoenix and 10 homes in Las Vegas that were sold recently.

Phoenix	258	269	229	279	249	260	242	240	307	
Las Vegas	280	245	319	289	259	268	295	239	262	250

Using the 5% level of significance, can you conclude that the current median price of homes in Phoenix is different from the current median price of homes in Las Vegas?

15.45 A factory's management is concerned about the number of defective parts produced by its machinists. Management suspects that production may be improved by giving machinists frequent breaks to reduce fatigue. Twenty-four randomly chosen machinists are randomly divided into two groups (A and B) of 12 each. During the next week all 24 machinists work to manufacture similar parts. The workers in Group A get a five-minute break every hour, whereas the workers in Group B stay on the usual schedule. The number of good parts produced by each machinist during the week is recorded in the following table.

Group A	157	139	188	143	172	144	191	128	177	160	175	162
Group B	160	118	150	165	158	159	127	133	170	164	152	142

At the 1% level of significance, can you conclude that the median number of good parts produced by machinists who take a five-minute break every hour is higher than the median number of good parts produced by the machinists who do not take a break?

15.46 Two brands of tires are tested to compare their durability. Eleven Brand X tires and 12 Brand Y tires are tested on a machine that simulates road conditions. The mileages (in thousands of miles for each tire) are shown in the following table.

Brand X	51	55	53	49	50.5	57	54.5	48.5	51.5	52	53.5	
Brand Y	48	47	54	55.5	50	51	46	49.5	52.5	51	49	45

Using the 5% level of significance, can you conclude that the median mileage for Brand X tires is greater than the median mileage for Brand Y tires?

15.47 Two Midwestern towns that are 120 miles apart are served by an airline that has been plagued by delays in the past few months. Consequently, many passengers who formerly traveled by air between these towns are taking advantage of a new express bus service. Some statistics students at a local college conducted a survey to see if the bus service between these towns was faster than the air flight. The students took random samples of 15 (one-way) plane trips and 17 (one-way) bus trips between the towns, recording the times for all 32 trips. The time recorded for each trip was measured from the scheduled departure time to the actual arrival time. The sum of the ranks for the 15 plane trips was 295; the sum of the ranks for the 17 bus trips was 233. At the 5% level of significance, can you conclude that the median time for the plane trip is higher than the median time for the bus trip?

15.4 The Kruskal-Wallis Test

In Chapter 12 we used the one-way analysis of variance (ANOVA) procedure to test whether or not the means of three or more populations are all equal. To apply the ANOVA procedure using the F distribution, we assumed that the populations from which the samples were drawn were normally distributed with equal variance, σ^2. However, if the populations being sampled are not normally distributed, then we cannot apply the ANOVA procedure of Chapter 12. In such cases, we can use the **Kruskal-Wallis test**, also called the Kruskal-Wallis H test. This is a nonparametric test because to use it we do not make any assumptions about the distributions of the populations being sampled. The only assumption we make is that all populations under consideration have identical shapes but differ only in location, which is measured by the median. Note that identical shapes do not mean that they have to have a normal distribution.

In a Kruskal-Wallis test, the null hypothesis is that the population distributions under consideration are all identical. The alternative hypothesis is that at least one of the population distributions differs and that, therefore, not all of the population distributions are identical. Note that we use the Kruskal-Wallis test to compare three or more populations. Also note that to apply the Kruskal-Wallis test, the size of each sample must be at least five.

Kruskal-Wallis Test To perform the Kruskal-Wallis test, we use the chi-square distribution that was discussed in Chapter 11. The test statistic in this test is denoted by H, which follows (approximately) the chi-square distribution. The critical value of H is obtained from Table VI in Appendix C for the given level of significance and $df = k - 1$, where k is the number of populations under consideration. *Note that the Kruskal-Wallis test is always right-tailed.*

To find the observed value of the test statistic H, we first rank the combined data from all samples in the same way as in a Wilcoxon rank sum test. The tied data values are handled the same way as in a Wilcoxon test. Then the observed value of H is calculated as explained below.

Observed Value of the Test Statistic *H* The observed value of the test statistic H is calculated using the following formula:

$$H = \frac{12}{n(n+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(n+1)$$

where

$$\begin{aligned}
R_1 &= \text{sum of the ranks for sample 1} \\
R_2 &= \text{sum of the ranks for sample 2} \\
&\vdots \\
R_k &= \text{sum of the ranks for sample } k \\
n_1 &= \text{sample size for sample 1} \\
n_2 &= \text{sample size for sample 2} \\
&\vdots \\
n_k &= \text{sample size for sample } k \\
n &= n_1 + n_2 + \cdots + n_k \\
k &= \text{number of samples}
\end{aligned}$$

The test statistic H measures the extent to which the k samples differ with regard to the ranks assigned to their data values. Basically, H is a measure of the variance of ranks (or of the variance of the means of ranks) for different samples. If all k samples have exactly the same mean of ranks, H will have a value of zero. The value of H becomes larger as the difference between the means of ranks for different samples increases. Thus, a larger observed value of H indicates that the distributions of the given populations do not seem to be identical.

Example 15–11 illustrates the procedure for applying the Kruskal-Wallis test.

■ EXAMPLE 15–11

Performing the Kruskal-Wallis test.

A researcher wanted to find out whether the population distributions of salaries of computer programmers are identical in three cities, Boston, San Francisco, and Atlanta. Three different samples—one from each city—produced the following data on the annual salaries (in thousands of dollars) of computer programmers.

Boston	San Francisco	Atlanta
43	54	57
39	33	68
62	58	60
73	38	44
51	43	39
46	55	28
	34	49
		57

Using the 2.5% significance level, can you conclude that the population distributions of salaries for computer programmers in these three cities are all identical?

Solution We apply the five steps to perform this hypothesis test.

Step 1. *State the null and alternative hypotheses.*

H_0: The population distributions of salaries of computer programmers in the three cities are all identical

H_1: The population distributions of salaries of computer programmers in the three cities are not all identical

Note that the alternative hypothesis states that the population distribution of at least one city is different from those of the other two cities.

Step 2. *Select the distribution to use.*

The shapes of the population distributions are unknown. We are comparing three populations. Hence, we apply the Kruskal-Wallis procedure to perform this test, and we use the chi-square distribution.

Step 3. *Determine the rejection and nonrejection regions.*

In this example,

$$\alpha = .025 \quad \text{and} \quad df = k - 1 = 3 - 1 = 2$$

Hence, from Table VI in Appendix C, the critical value of χ^2 is 7.378, as shown in Figure 15.11.

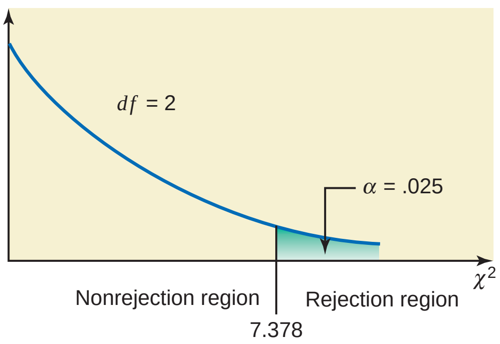


Figure 15.11

Step 4. *Calculate the value of the test statistic.*

To calculate the observed value of the test statistic H, we first rank the combined data for all three samples and find the sum of ranks for each sample separately. This is done in Table 15.7.

Table 15.7

Boston		San Francisco		Atlanta	
Salary	**Rank**	**Salary**	**Rank**	**Salary**	**Rank**
43	7.5	54	13	57	15.5
39	5.5	33	2	68	20
62	19	58	17	60	18
73	21	38	4	44	9
51	12	43	7.5	39	5.5
46	10	55	14	28	1
		34	3	49	11
				57	15.5
$n_1 = 6$	$R_1 = 75$	$n_2 = 7$	$R_2 = 60.5$	$n_3 = 8$	$R_3 = 95.5$

We have

$$n = n_1 + n_2 + n_3 = 6 + 7 + 8 = 21$$

and

$$H = \frac{12}{n(n+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(n+1)$$

$$= \frac{12}{21(21+1)}\left(\frac{(75)^2}{6} + \frac{(60.5)^2}{7} + \frac{(95.5)^2}{8}\right) - 3(21+1)$$

$$= 1.543$$

Step 5. *Make a decision.*

Because the observed value of $H = 1.543$ is less than the critical value of $H = 7.378$ and it falls in the nonrejection region, we do not reject the null hypothesis. Consequently, we conclude that the population distributions of salaries of computer programmers in the three cities seem to be all identical. ■

EXERCISES

CONCEPTS AND PROCEDURES

15.48 Briefly explain when the Kruskal-Wallis test is used to make a test of hypothesis.

15.49 What assumption that is required for the ANOVA procedure of Chapter 12 is not necessary for the Kruskal-Wallis test?

15.50 Describe the form of the null and alternative hypotheses for a Kruskal-Wallis test.

15.51 Here, n_i is the size of the ith sample and R_i is the sum of ranks for the ith sample. For each of the following cases, perform the Kruskal-Wallis test using the 5% level of significance.

a. $n_1 = 9$, $n_2 = 8$, $n_3 = 5$, $R_1 = 81$, $R_2 = 102$, $R_3 = 70$
b. $n_1 = n_2 = n_3 = n_4 = 5$, $R_1 = 27$, $R_2 = 30$, $R_3 = 83$, $R_4 = 70$
c. $n_1 = 6$, $n_2 = 10$, $n_3 = 6$, $R_1 = 93$, $R_2 = 70$, $R_3 = 90$
d. $n_1 = 8$, $n_2 = 9$, $n_3 = 8$, $n_4 = 10$, $n_5 = 9$,
$R_1 = 210$, $R_2 = 195$, $R_3 = 178$, $R_4 = 212$, $R_5 = 195$

15.52 The following table gives the ranked data for three samples. Perform the Kruskal-Wallis test using the 1% level of significance.

Sample I	Sample II	Sample III
3	14	2
1	11	4.5
10	16	13
7	15	4.5
9	12	8
6		

APPLICATIONS

15.53 Refer to Examples 12–2 and 12–3 of Chapter 12. Fifteen randomly selected fourth-grade students were randomly assigned to three groups, and each group was taught arithmetic by a different method. At the end of the semester, all 15 students took the same arithmetic test. Their test scores are given in the following table.

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65	69	74
87	90	67

a. At the 1% level of significance, can you reject the null hypothesis that the median arithmetic test scores of all fourth-grade students taught by these three methods are all equal?
b. Compare your answer to part a with the result of the hypothesis test in Example 12–3.

15.54 A consumer agency investigated the premiums charged by four auto insurance companies. The agency randomly selected five drivers insured by each company who had similar driving records, autos, and insurance coverages. The following table gives the monthly premiums paid by the 20 drivers.

Company A	Company B	Company C	Company D
$65	$48	$57	$62
73	69	61	53
54	88	89	45
43	75	77	51
70	72	69	44

Can you reject the null hypothesis that the distributions of auto insurance premiums paid per month by all such drivers are the same for all four companies? Use $\alpha = .05$.

15.55 Refer to Problem 10 of the Self-Review Test in Chapter 12. A small college town has four pizza parlors that make deliveries. A student doing a research paper for her business management class decides to compare how promptly the four parlors deliver. On six randomly chosen nights, she orders a large pepperoni pizza from each establishment and then records the elapsed time until the pizza is delivered to her apartment. Assume that her apartment is approximately at the same distance from the four pizza parlors. The following table shows the delivery times (in minutes) for these orders.

Tony's	Luigi's	Angelo's	Kowalski's
20.0	22.1	22.3	23.9
24.0	27.0	26.0	24.1
18.3	20.2	24.0	25.8
22.0	32.0	30.1	29.0
20.8	26.0	28.0	25.0
19.0	24.8	25.8	24.2

a. Test the null hypothesis that the distributions of delivery times are identical for the four pizza parlors. Use the 5% level of significance.
b. Compare your conclusion of part a here with that of part a of Problem 10 of the Self-Review Test in Chapter 12.

15.56 Refer to Exercise 12.27 of Chapter 12. A resort area has three seafood restaurants, which employ students during the summer season. The local chamber of commerce took a random sample of five servers from each restaurant and recorded the tips they received on a recent Friday night. The results of the survey are shown in the table below. Assume that the Friday night for which the data were collected is typical of all Friday nights of the summer season.

Barzini's	Hwang's	Jack's
$ 97	$67	$ 93
114	85	102
105	92	98
85	78	80
120	90	91

a. Would a student seeking a server's job at one of these three restaurants conclude that the population distributions of tips on Friday nights are identical for the three restaurants? Use the 5% level of significance.
b. Compare your conclusion of part a with that of part a of Exercise 12.27 of Chapter 12.
c. What would your decision be if the probability of making a Type I error were zero in part a? Explain.

15.57 A factory operates three shifts a day, five days per week, each with the same number of workers and approximately the same level of production. The following table gives the number of defective parts produced during each shift over a period of five days.

First Shift	Second Shift	Third Shift
23	25	33
36	35	44
32	41	50
40	38	52
45	50	60

At the 5% level of significance, can you conclude that the median number of defective parts is the same for all three shifts?

15.58 A consumer group wanted to compare the service time at three fast-food restaurants, Al's, Eduardo's, and Patel's. Every Tuesday and Wednesday for four weeks, three staff members of the group were randomly assigned to these three restaurants. Each staff member went to his or her assigned restaurant and ordered a hamburger, fries, and a Coke and then recorded the time that elapsed from entering the restaurant until receiving the food. The service times (in minutes) for these eight days for the three restaurants are listed below.

Al's	Eduardo's	Patel's
7.0	3.3	1.1
8.3	11.0	2.4
6.9	5.7	1.8
1.3	8.1	3.0
6.7	6.6	4.1
7.1	13.0	12.0
5.5	2.3	1.5
6.6	5.9	3.1

Assume that these service times make up random samples of all service times at the respective restaurants. At the 10% level of significance, can you conclude that there is a difference in the median service times at these three restaurants?

15.5 The Spearman Rho Rank Correlation Coefficient Test

In Chapter 13 we discussed the linear correlation coefficient between two variables x and y. We also learned how to make a test of hypothesis about the population correlation coefficient ρ using the information from a sample. In that chapter we used the t distribution to perform this test about ρ. However, using the procedure of Chapter 13 and using the t distribution to make this test of hypothesis about ρ require that both variables x and y are normally distributed.

The **Spearman rho rank correlation coefficient** (Spearman's rho) is a nonparametric analog of the linear correlation coefficient of Chapter 13. It helps us decide what type of relationship, if any, exists between data from populations with unknown distributions. The Spearman rho rank correlation coefficient is denoted by r_s for sample data and by ρ_s for population data. This correlation coefficient is simply the linear correlation coefficient between the ranks of the data on variables x and y. To make a test of hypothesis about the Spearman rho rank correlation coefficient, we do not need to make any assumptions about the populations of x and y variables.

Spearman Rho Rank Correlation Coefficient The Spearman rho rank correlation coefficient is denoted by r_s for sample data and by ρ_s for population data. This correlation coefficient is simply the linear correlation coefficient between the ranks of the data. To calculate the value of r_s, we rank the data for each variable, x and y, separately and denote those ranks by u and v, respectively. Then we take the difference between each pair of ranks and denote it by d. Thus,

$$\text{Difference between each pair of ranks} = d = u - v$$

Next, we square each difference d and add these squared differences to find Σd^2. Finally, we calculate the value of r_s using the formula:

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

In a test of hypothesis about the Spearman rho rank correlation coefficient ρ_s, the test statistic is r_s and its observed value is calculated by using the above formula.

Example 15–12 shows how to calculate the Spearman rho rank correlation coefficient r_s and how to perform a test of hypothesis about ρ_s.

■ EXAMPLE 15–12

Performing the Spearman rho rank correlation coefficient test.

Suppose we want to investigate the relationship between the per capita income (in thousands of dollars) and the infant mortality rate (in percent) for different states. The following table gives data on these two variables for a random sample of eight states.

Per capita income (x)	29.85	19.0	19.18	31.78	25.22	16.68	23.98	26.33
Infant mortality (y)	8.3	10.1	10.3	7.1	9.9	11.5	8.7	9.8

Based on these data, can you conclude that there is no significant (linear) correlation between the per capita incomes and the infant mortality rates for all states? Use $\alpha = .05$.

Solution We perform the five steps to test the null hypothesis that there is no correlation between the two variables against the alternative hypothesis that there is a significant correlation.

Step 1. *State the null and alternative hypotheses.*

The null and alternative hypotheses are as follows:

H_0: There is no correlation between per capita incomes and infant mortality rates in all states

H_1: There is a correlation between per capita incomes and infant mortality rates in all states

If we denote the Spearman correlation coefficient by ρ_s, the null hypothesis and the alternative hypothesis can be written as

$$H_0: \rho_s = 0$$

$$H_1: \rho_s \neq 0$$

Note that this is a two-tailed test.

Step 2. *Select the distribution to use.*

Because the sample is taken from a small population and the variables do not follow a normal distribution, we use the Spearman rho rank correlation coefficient test procedure to make this test.

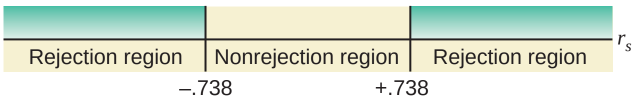
Step 3. *Determine the rejection and nonrejection regions.*

The test statistic that is used to make this test is r_s, and its critical values are given in Table XI that appears at the end of this chapter. Note that, for this example,

$$n = 8 \quad \text{and} \quad \alpha = .05$$

To read the critical value of r_s from Table XI, we locate 8 in the column labeled n and .05 in the top row of the table for a two-tailed test. The critical values of r_s are $\pm.738$, or $+.738$ and $-.738$. Thus, we will reject the null hypothesis if the observed value of r_s is either $-.738$ or less, or $+.738$ or greater. The rejection and nonrejection regions for this example are shown in Figure 15.12.

Figure 15.12



Critical Value of r_s The critical value of r_s is obtained from Table XI for the given sample size and significance level. If the test is two-tailed, we use two critical values, one negative and one positive. However, we use only the negative value of r_s if the test is left-tailed, and only the positive value of r_s if the test is right-tailed.

Step 4. *Calculate the value of the test statistic.*

In the Spearman rho rank correlation coefficient test, the test statistic is denoted by r_s, which is simply the linear correlation coefficient between the ranks of the data. As explained in the beginning of this section, to calculate the observed value of r_s, we use the formula:

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

where $d = u - v$, and u and v are the ranks of variables x and y, respectively.

Table 15.8

u	7	2	3	8	5	1	4	6	
v	2	6	7	1	5	8	3	4	
d	5	-4	-4	7	0	-7	1	2	
d^2	25	16	16	49	0	49	1	4	$\Sigma d^2 = 160$

Table 15.8 shows the ranks for x and y, which are denoted by u and v, respectively. The table also lists the values of d, d^2, and Σd^2. Note that if two or more values are equal, we use the average of their ranks for all of them. Hence, the observed value of r_s is

$$r_s = 1 - \frac{6(160)}{8(64 - 1)} = 1 - \frac{960}{504} = -.905$$

Note that Spearman's rho rank correlation coefficient has the same properties as the linear correlation coefficient (discussed in Chapter 13). Thus, $-1 \le r_s \le 1$ or $-1 \le \rho_s \le 1$, depending on whether sample or population data are used to calculate the Spearman rho rank correlation coefficient. If $\rho_s = 0$, there is no relationship between the x and y data. If $0 < \rho_s \le 1$, on average a larger value of x is associated with a larger value of y. Similarly, if $-1 \le \rho_s < 0$, on average a larger value of x is associated with a smaller value of y.

Step 5. *Make a decision.*

Because $r_s = -.905$ is less than $-.738$ and it falls in the rejection region, we reject H_0 and conclude that there is a correlation between the per capita incomes and the infant mortality rates in all states. Because the value of r_s from the sample is negative, we can also state that as per capita income increases, infant mortality tends to decrease. ■

Decision Rule for the Spearman Rho Rank Correlation Coefficient The null hypothesis is always H_0: $\rho_s = 0$. The observed value of the test statistic is always the value of r_s computed from the sample data. Let α denote the significance level, and $-c$ and $+c$ be the critical values for the Spearman rho rank correlation coefficient test obtained from Table XI.

1. For a two-tailed test, the alternative hypothesis is H_1: $\rho_s \ne 0$. If $\pm c$ are the critical values corresponding to sample size n and two-tailed α, we reject H_0 if either $r_s \le -c$ or $r_s \ge +c$; that is, reject H_0 if r_s is "too small" or "too large."
2. For a right-tailed test, the alternative hypothesis is H_1: $\rho_s > 0$. If $+c$ is the critical value corresponding to sample size n and one-sided α, we reject H_0 if $r_s \ge +c$; that is, reject H_0 if r_s is "too large."
3. For a left-tailed test, the alternative hypothesis is H_1: $\rho_s < 0$. If $-c$ is the critical value corresponding to sample size n and one-sided α, we reject H_0 if $r_s \le -c$; that is, reject H_0 if r_s is "too small."

EXERCISES

CONCEPTS AND PROCEDURES

15.59 What assumptions that are required for hypothesis tests about the linear correlation coefficient ρ in Chapter 13 are not required for testing a hypothesis about the Spearman rho rank correlation coefficient?

15.60 Two sets of paired data on two variables, x and y, have been ranked. In each case, the ranks for x and y are denoted by u and v, respectively, and are shown in the tables. Calculate the Spearman rho rank correlation coefficient for each case.

a.

u	2	1	3	4	6	5	7	8
v	8	6	7	4	5	2	1	3

b.

u	1	2	3	4	5	6	7
v	4	2	1	5	3	7	6

15.61 Calculate the Spearman rho rank correlation coefficient for each of the following data sets.

a.

x	5	10	15	20	25	30
y	17	15	12	14	10	9

b.

x	27	15	32	21	16	40	8
y	95	81	102	88	75	120	62

15.62 Perform the indicated hypothesis test in each of the following cases.

a. $n = 9$, $r_s = .575$, H_0: $\rho_s = 0$, H_1: $\rho_s > 0$, $\alpha = .025$
b. $n = 15$, $r_s = -.575$, H_0: $\rho_s = 0$, H_1: $\rho_s < 0$, $\alpha = .005$
c. $n = 20$, $r_s = .554$, H_0: $\rho_s = 0$, H_1: $\rho_s \neq 0$, $\alpha = .01$
d. $n = 20$, $r_s = .554$, H_0: $\rho_s = 0$, H_1: $\rho_s > 0$, $\alpha = .01$

APPLICATIONS

15.63 The following data are a random sample of the heights (in inches) and weights (in pounds) of 10 NBA players selected at random.

Height	84	76	79	79	84	74	83	81	83	75
Weight	240	208	205	215	265	182	225	220	250	190

a. Based on the reasonable assumption that as height increases, weight tends to increase, do you expect the value of r_s to be positive or negative? Why?
b. Compute the value of r_s. Does it agree with your expectation of its value in part a?

15.64 Let ρ_s be the Spearman rho rank correlation coefficient between heights (in inches) and weights (in pounds) for the entire population of NBA players listed in Data Set II. Using the value of r_s computed from the sample data in Exercise 15.63, test the null hypothesis H_0: $\rho_s = 0$ against the alternative hypothesis H_1: $\rho_s > 0$ at the significance level of $\alpha = .01$.

15.65 In Example 13–1 of Chapter 13, we estimated the regression line for the data given in Table 13.2 on food expenditures (in hundred dollars) and incomes (in hundred dollars). Those data are reproduced here.

Income (x)	35	49	21	39	15	28	25
Food expenditure (y)	9	15	7	11	5	8	9

The estimated regression line in Example 13–1 had a slope of .2642.
a. Do you expect the Spearman rho rank correlation coefficient for these data to be positive or negative? Why?
b. Compute r_s for these data. Did it come out as expected?

15.66 In Example 13–7 of Chapter 13, the null hypothesis H_0: $\rho = 0$ was tested against the alternative hypothesis H_1: $\rho > 0$, where ρ is the linear correlation coefficient for the population. At a significance level of $\alpha = .01$, H_0: $\rho = 0$ was rejected there. Hence, it seemed that in fact $\rho > 0$.
a. If ρ_s is the Spearman rho rank correlation coefficient for the entire population for food expenditure and income data, and you test H_0: $\rho_s = 0$ against H_1: $\rho_s > 0$ at the significance level of $\alpha = .01$, based on the results of Example 13–7, do you expect to reject or accept H_0? Why?
b. Perform the hypothesis test stated in part a.

15.67 The following table shows the combined (math and verbal) SAT scores (denoted by x) and college grade point averages (denoted by y) on completion of a bachelor's degree for nine randomly chosen recent college graduates who had taken the SAT test.

x	1105	990	1040	1215	1405	975	1300	1010	1080
y	3.33	2.62	3.05	3.60	3.85	2.43	3.90	2.40	2.95

a. Find the Spearman rho rank correlation coefficient for this data set.
b. Test H_0: $\rho_s = 0$ against H_1: $\rho_s > 0$ using the 5% level of significance.
c. Does your test result indicate a positive relationship between the variables x and y?

15.68 A day-care center operator is concerned about the aggressive behavior of young boys left in her care. She feels that prolonged watching of television tends to promote aggressive behavior. She selects seven boys at random and ranks them according to the level of aggressiveness in their behavior, with a rank of 7 indicating most aggressive and a rank of 1 denoting least aggressive. Then she asks each boy's parent(s) to estimate the average number of hours per week the boy spends watching television. The following table shows the data collected on the aggressiveness ranks of these boys and the number of hours spent watching TV per week.

Aggressiveness rank	1	2	3	4	5	6	7
Weekly TV hours	15	21	28	8	24	32	20

a. Calculate r_s for these data.
b. At the 5% level of significance, can you conclude that there is a positive relationship between aggressiveness and hours spent watching television?

15.6 The Runs Test for Randomness

The **runs test for randomness** tests the null hypothesis that a sequence of events has occurred randomly against the alternative hypothesis that this sequence of events has not occurred randomly.

The Small-Sample Case

As an example of a runs test, in apartment complexes families with children are often assigned to units close to one another to lessen the impact of noise on childless tenants. We would like to determine whether a landlord has randomly assigned units to tenants irrespective of whether they have children, or has tried to cluster tenants with children. Here what we really mean by *random* is independence in the sense that if, say, a childless tenant lives in unit 3, this provides no additional information about whether the tenants in units 2 and 4 are childless. With non-randomness, we would have some additional information.

Suppose that a part of the apartment complex consists of a single building with 10 adjacent units numbered 1 to 10. We specify the family status of the 10 tenants via a string of 10 letters, using C for "has children" and D for "no children." One possible string is D D C D C C D D C C, which would mean that the tenants in units 3, 5, 6, 9, and 10 have children and those in 1, 2, 4, 7, and 8 do not. (Note that in our example the numbers of tenants with and without children are equal, but they do not have to be equal.)

Two extreme arrangements that may not be random are C C C C C D D D D D and C D C D C D C D C D. In the first arrangement, all the tenants with children are adjacent, as are those without children. In the second case, tenants with and without children alternate perfectly. Note that in these two examples exactly half of the tenants have children and half do not.

A characteristic of a string of the letters C and D that helps determine the randomness of the string is called a **run**. A run is a sequence of the same symbol (in this case the letter C or D) appearing one or more times. The arrangement C C C C C D D D D D has two runs; the arrangement C D C D C D C D C D has 10 runs. Intuitively, we know that in the string with 2 runs the arrangement is not random because there are too few runs, whereas in the string with 10 runs the arrangement is not random due to too many runs.

If the number of tenants with children, the number without children, and their arrangement in the 10 units are all random, then the number of runs in the arrangement, which we denote by $\boldsymbol{R}$, will also be random. Thus, R is a statistic with its own sampling distribution. Table XII (that appears at the end of this chapter) gives the critical values of R for a significance level of 5%—that is, for $\alpha = .05$. There are two parameters associated with the distribution of R, n_1 and n_2. Here n_1 is the number of times the first symbol (in our example the letter C) appears in the string, and n_2 is the number of times the second symbol (in our example the letter D) appears. Table XII provides critical values of R for values of n_1 and n_2 up to 15. If either $n_1 > 15$ or $n_2 > 15$, we can apply the normal approximation (discussed later in the section on the large-sample case) to perform the test. For each pair of n_1 and n_2, there are two critical values: a smaller value (denoted by c_1) and a larger value (denoted by c_2).

Definition

Run A *run* is a sequence of one or more consecutive occurrences of the same outcome in a sequence of occurrences in which there are only two outcomes. The number of runs in a sequence is denoted by R. The value of R obtained for a sequence of outcomes for a sample gives the observed value of the test statistic for the runs test for randomness.

Suppose we formally set up the following hypotheses:

H_0: Tenants with and without children are randomly mixed among the 10 units

H_1: These tenants are not randomly mixed

We will reject H_0 if either of the following occurs:

$$R \le c_1 \text{(too few runs)} \quad \text{or} \quad R \ge c_2 \text{(too many runs)}$$

Let's apply these rules to the hypothetical strings listed earlier to determine whether or not to reject H_0 at a significance level of $\alpha = .05$.

1. Let the string of letters be C C C C C D D D D D. Here $n_1 = 5$, $n_2 = 5$, and $R = 2$. From Table XII, $c_1 = 2$ and $c_2 = 10$. Since $R \le c_1$, we reject H_0 on the basis that there are too few runs.
2. Let the string of letters be C D C D C D C D C D. Here $n_1 = 5$, $n_2 = 5$, and $R = 10$. From Table XII, $c_1 = 2$ and $c_2 = 10$. Since $R \ge c_2$, we reject H_0 on the basis that there are too many runs.
3. Let the string of letters be D D C D C C D D C C. Here $n_1 = 5$, $n_2 = 5$, and $R = 6$. From Table XII, $c_1 = 2$ and $c_2 = 10$. Because the value of $R = 6$ is between $c_1 = 2$ and $c_2 = 10$, we do not reject H_0.

Example 15–13 illustrates the application of the runs test for randomness.

■ EXAMPLE 15–13

Performing the runs test for randomness: small sample.

A college admissions office is interested in knowing whether applications for admission arrive randomly with respect to gender. The genders of 25 consecutively arriving applications were found to arrive in the following order (here M denotes a male applicant and F a female applicant).

M F M M F F F M F M M M F F F F M M M F F M F M M

Can you conclude that the applications for admission arrive randomly with respect to gender? Use $\alpha = .05$.

Solution We perform the following five steps in this hypothesis test.

Step 1. *State the null and alternative hypotheses.*

H_0: Applications arrive in a random order with respect to gender

H_1: Applications do not arrive in a random order with respect to gender

Step 2. *Select the distribution to use.*

Let n_1 and n_2 be the number of male and female applicants, respectively. Then

$$n_1 = 13 \quad \text{and} \quad n_2 = 12$$

Because both n_1 and n_2 are less than 15, we use the runs test to check for randomness.

Step 3. *Determine the rejection and nonrejection regions.*

For $n_1 = 13$, $n_2 = 12$, and $\alpha = .05$, the critical values from Table XII are $c_1 = 8$ and $c_2 = 19$. Thus, we will not reject the null hypothesis if the observed value of R is in the interval 9 to 18. We will reject the null hypothesis if the observed value of R is either 8 or lower, or 19 or higher. The rejection and nonrejection regions are shown in Figure 15.13.

Figure 15.13

8 or lower	9 to 18	19 or higher	R
Rejection region	Nonrejection region	Rejection region	

Step 4. *Calculate the value of the test statistic.*

As the given data show, the 25 applications included in the sample were received in the following order with respect to gender:

M F M M F F F M F M M M F F F F M M M F F M F M M

Because this string of the letters M and F has 13 runs,

$$\text{Observed value of } R = 13$$

Step 5. *Make a decision.*

Because $R = 13$ is between 9 and 18, we do not reject H_0. Hence, we conclude that the applications for admission arrive in a random order with respect to gender. ■

The Large-Sample Case

If *either* $n_1 > 15$ *or* $n_2 > 15$, the sample is considered large for the purpose of applying the runs test for randomness and we use the normal distribution to perform the test.

Observed Value of z For large values of n_1 and n_2, the distribution of R (the number of runs in the sample) is approximately normal with its mean and standard deviation given as

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad \text{and} \quad \sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

The observed value of z for R is calculated using the formula

$$z = \frac{R - \mu_R}{\sigma_R}$$

In this case, rather than using Table XII to find the critical values of R, we use the standard normal distribution table (Table IV in Appendix C) to find the critical values of z for the given significance level. Then, we make a decision to reject or not to reject the null hypothesis based on whether the observed value of z falls in the rejection or the nonrejection region. Example 15–14 describes the application of this procedure.

■ EXAMPLE 15–14

Performing the runs test for randomness: large sample.

Refer to Example 15–13. Suppose that the admissions officer examines 50 consecutive applications and observes that $n_1 = 22$, $n_2 = 28$, and $R = 20$, where n_1 is the number of male applicants, n_2 the number of female applicants, and R the number of runs. Can we conclude that the applications for admission arrive randomly with respect to gender? Use $\alpha = .01$.

Solution We perform the following five steps to make this test.

Step 1. *State the null and alternative hypotheses.*

H_0: Applications arrive in a random order with respect to gender

H_1: Applications do not arrive in a random order with respect to gender

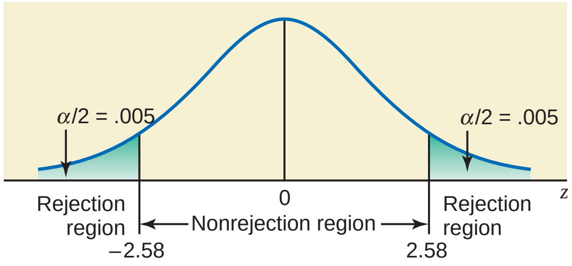
Step 2. *Select the distribution to use.*

Here, $n_1 = 22$ and $n_2 = 28$. Since n_1 and n_2 are both greater than 15, we use the normal distribution to make the runs test. Note that only one of n_1 and n_2 has to be greater than 15 to apply the normal distribution.

Step 3. *Determine the rejection and nonrejection regions.*

The significance level is .01 and the test is two-tailed. From Table IV in Appendix C, the critical values of z for .005 and .9950 areas to the left are -2.58 and 2.58, respectively. The rejection and nonrejection regions are shown in Figure 15.14.

Figure 15.14



Step 4. *Calculate the value of the test statistic.*

To find the observed value of z, we first find the mean and standard deviation of R as follows:

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{(2)(22)(28)}{22 + 28} + 1 = 25.64$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} = \sqrt{\frac{2(22)(28)(2 \cdot 22 \cdot 28 - 22 - 28)}{(22 + 28)^2(22 + 28 - 1)}} = 3.44783162$$

The observed value of the test statistic z is

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{20 - 25.64}{3.44783162} = -1.64$$

Note that the value of R is given in the example to be 20.

Step 5. *Make a decision.*

Since $z = -1.64$ is between -2.58 and 2.58, we do not reject H_0, and we conclude that the applications for admission arrive in a random order with respect to gender. ■

EXERCISES

CONCEPTS AND PROCEDURES

15.69 Briefly explain the term *run* as used in a runs test.

15.70 What is the usual form of the null hypothesis in a runs test for randomness?

15.71 Under what conditions may we use the normal distribution to perform a runs test?

15.72 Using the runs test for randomness, indicate whether the null hypothesis should be rejected in each of the following cases.

a. $n_1 = 10$, $n_2 = 12$, $R = 17$, $\alpha = .05$
b. $n_1 = 20$, $n_2 = 23$, $R = 35$, $\alpha = .01$
c. $n_1 = 15$, $n_2 = 17$, $R = 7$, $\alpha = .05$
d. $n_1 = 14$, $n_2 = 13$, $R = 21$, $\alpha = .05$

15.73 In Example 15–13, if we use the symbol 0 for male and 1 for female instead of M and F, would this affect the test in any way? Why or why not?

15.74 For each of the following sequences of observations, determine the values of n_1, n_2, and R.

a. X X Y X Y Y X Y X Y X X X Y Y
b. F M F F F F M M F F F F F F
c. + + + − − − − − − − + + − + − + − + + + + + + +
d. 1 1 0 0 0 0 1 1 0 0 1 1 1 1

APPLICATIONS

15.75 A psychic claims that she can cause a nonrandom sequence of heads (H) and tails (T) to appear when a coin is tossed a number of times. A fair coin was tossed 20 times in her presence, and the following sequence of heads and tails was obtained.

H H T H T H T T H T H H T T H H H T T H

Can you conclude that the psychic's claim is true? Use $\alpha = .05$.

15.76 At a small soda factory, the amount of soda put into each 12-ounce bottle by the bottling machine varies slightly. The plant manager suspects that the machine has a nonrandom pattern of overfilling and underfilling the bottles. The following are the results of filling 18 bottles, where O denotes 12 ounces or more of soda in a bottle and U denotes less than 12 ounces of soda.

U U U O O O O U U O O O U O U U U U

Using the runs test at the 5% significance level, can you conclude that there is a nonrandom pattern of overfilling and underfilling such bottles?

15.77 An experimental planting of a new variety of pear trees consists of a single row of 20 trees. Several of these trees were affected by an unknown disease. The order of diseased and normal trees is shown below, where D denotes a diseased and N denotes a normal tree.

N N N D D D N N N N N N D D D D N N N N

If the sequence of diseased and normal trees falls into a nonrandom pattern with clusters of diseased trees, that would suggest that the disease may be contagious. Perform the runs test at the 5% significance level to determine if there is evidence of a nonrandom pattern in the sequence.

15.78 A fourth-grade teacher asks her class orally one by one whether "potato" or "potatoe" is the correct spelling for the common vegetable. She suspects the children may engage in copycat behavior, in which an incorrect spelling by one student is likely to be followed by an incorrect spelling by the next student. If this theory is true, there should be fewer runs than expected. The responses of the 22 students in the class are as follows, where a correct answer is denoted by C and an incorrect answer by I.

C C C I I I C C C C C I I C C C C I C C C C

At the 5% significance level, test the null hypothesis that the correct (and incorrect) responses are randomly distributed in the population of all fourth-graders against the alternative hypothesis that they are not randomly distributed in the population. Assume that this class is a random sample of all fourth-graders.

15.79 Do baseball players' hits come in streaks? Seventy-five consecutive at-bats of a baseball player were recorded to determine whether the hits were randomly distributed or nonrandomly distributed for him, thus possibly indicating the presence of "hitting streaks." The observation of these 75 at-bats produced the following data.

$$n_1 = \text{number of hits} = 22, \quad n_2 = \text{number of nonhits} = 53, \quad R = \text{number of runs} = 37$$

Can you conclude that hits occur randomly for this player? Use $\alpha = .01$.

15.80 A researcher wanted to determine whether the stock market moves up and down randomly. He recorded the movement of the Dow Jones Industrial Average for 40 consecutive business days. The observed data showed that the market moved up 16 times, moved down 24 times, and had 11 runs during these 40 days. Using a significance level of 5%, do you think that the movements in the stock market are random?

15.81 Many state lotteries offer a "daily number" game, in which a three-digit number is randomly drawn every day. Suppose that a certain state generates its daily number by a computer program, and the state lottery commissioner suspects that the process is flawed. Specifically, she feels that if today's lucky number is higher than 500, tomorrow's number is more likely to exceed 500; and if today's number is below 500, tomorrow's number is more likely to be under 500. Suppose that a sequence of the state's lucky numbers for a period of consecutive days is analyzed for runs above 500 and runs below 500. If the commissioner is right, there should be fewer runs than expected by chance. Analysis of a sequence for 50 consecutive days produced the following information.

$$n_1 = 27 \text{ numbers above } 500 \quad n_2 = 23 \text{ numbers below } 500 \quad R = \text{number of runs} = 11$$

Using the 2.5% level of significance, can you conclude that the sequence of all this state's daily numbers for this game is nonrandom?

USES AND MISUSES...

1. I'M FREE!

Imagine that you turn on the news one morning. Still a bit sleepy, you hear a report: "And this just in from the state government: all road signs have been removed. That's right, folks. No more traffic lights, no more 'no parking' signs, no more speed limits! And now, the weather… " You would rightfully be scared to death. During your drive to school or work, you would be much more vigilant than the day before. You would approach an intersection more slowly, you would have to rely on published maps given the absence of exit signs, and you would keep your eye out for very fast and very dangerous drivers. Quite simply, you would not be able to rely on assumptions regarding the rules of the road.

Assumptions can be good. This text has emphasized that the parametric methods (including hypothesis testing, chi-square tests, *F*-tests, and linear regression models), require that your samples, populations, and errors be normally distributed. Because of the central limit theorem, the assumption that population data and large samples taken from them are normally distributed is typically pretty good. Standard statistical computer packages are able to compare your samples to a normal distribution and should be used to do so. The assumption that a statistical sample comes from a normal distribution is equivalent to augmenting your data set. In the example above, the assumption is roughly equivalent to the rules of the road: You do not have to check on the driver next to you all the time because you can assume that he or she will behave in a certain way.

However, for every case in which the assumption of normality applies, there is at least one in which it does not. Without the distribution parameters of the mean and variance, you are going to need to gather more data to make up for the fact that you do not have these parameters, or accept a larger confidence interval than you would have preferred. You might even find that as you collect more data, they resemble a normal distribution after all. Additionally, the nonparametric tests have their own applications and assumptions, and keeping them straight can be tricky: The Kruskal-Wallis test investigates if the distributions under investigation are identical but tells us no more.

Fortunately, the nonparametric methods do not abandon the rules of the road. Each nonparametric method is based on properties of probability distributions and the assumption of random sampling of the populations under investigation. Each has explicit null and alternative hypotheses, and proper application requires a detailed specification. Similarly, if you were to drive your car down the road on that dangerous morning, you would probably find that everyone was still driving on the right side of the road.

2. MORE POWER TO YOU!

Chapters 8 through 10, and 12 through 14 discussed a variety of (parametric) procedures related to statistical inference and each of these procedures requires that certain conditions hold true in order to produce valid results. The nonparametric methods presented in Chapter 11 and 15 also require a set of conditions that must hold true, but these conditions are much less restrictive than the ones in other chapters. If the results of a parametric test are valid, then the results from an equivalent nonparametric test will also be valid. However, the inverse of this relationship is not true. Having said that, many people wonder why not just forget about the stricter conditions and instead use a nonparametric test each time? That is a good question, and it has a good answer.

As you may recall from Chapter 9, every hypothesis test has the potential to produce an incorrect result. The possible incorrect results were referred to as Type I and Type II errors in that chapter. You may remember from Chapter 9 that the significance level (denoted by α) gives the probability of a Type I error. In practice, the statistician specifies the significance level that is used in a hypothesis test. The probability of a Type II error (denoted by β) in a hypothesis test problem is a function of the test being used, the sample size, and the significance level. The *power* of a hypothesis test is defined to be 1 minus the probability of a Type II error, i.e., $1 - \beta$. So we try to use a test that has the highest power of the (hypothesis) test given a specific significance level.

Parametric tests, such as the single-sample *t* test for the mean and a one-way ANOVA test, have higher powers of the test than their nonparametric compatriots. Therefore, if it is reasonable to conclude that the stricter conditions of a parametric test have been met, using a parametric test will result in a higher power of the test. However, if it is not reasonable to assume that the conditions of a parametric test have been met, then using a parametric test will not produce valid results. So while the power of the test will be lower when using a nonparametric test, you can rely on the fact that the designated significance level will be met.

Glossary

Categorical data Data divided into different categories for identification purposes only.

Distribution-free test A hypothesis test in which no assumptions are made about the specific population distribution from which the sample is selected.

H The test statistic used in the Kruskal-Wallis test.

Kruskal-Wallis test A distribution-free method used to test the hypothesis that three or more populations have identical distributions.

Nonparametric test A hypothesis test in which the sample data are not assumed to come from a specific type of population distribution, such as the normal distribution.

ρ_s The value of the Spearman rho rank correlation coefficient between the ranks of the values of two variables for population data.

r_s The value of the Spearman rho rank correlation coefficient between the ranks of the values of two variables for sample data.

R The number of runs in a runs test used to test for randomness.

Run A sequence of one or more consecutive occurrences of the same outcome in a sequence of occurrences in which there are only two possible outcomes.

Runs test for randomness A test that is used to test the null hypothesis that a sequence of events has occurred randomly.

Sign test A nonparametric test that is used to test a population proportion (with categorical data), a population median (with numerical data), or the difference in population medians for two dependent and paired sets of numerical data.

Spearman rho rank correlation coefficient The linear correlation coefficient between the ranks of paired data for two samples or populations.

T The test statistic used in the Wilcoxon signed-rank test and Wilcoxon rank sum test.

T_U, T_L The upper and lower critical values for the Wilcoxon rank sum test obtained from the table.

Wilcoxon rank sum test A nonparametric test that is used to test whether two independent samples come from identically distributed populations by analyzing the ranks of the pooled sample data.

Wilcoxon signed-rank test A nonparametric test that is used to test whether two paired and dependent samples come from identically distributed populations by analyzing the ranks of the paired differences of the samples.

Supplementary Exercises

15.82 Fifteen cola drinkers are given two paper cups, one containing Brand A cola and the other containing Brand B cola. Each person tries both drinks and then indicates which one he or she prefers. The drinks are offered in random order (some people are given Brand A first, and others get Brand B first). Ten of the people prefer Brand A, while five prefer Brand B. Using the sign test at the .05 significance level, can you conclude that among all cola drinkers there is a preference for Brand A?

15.83 Twenty-four randomly selected people are given samples of two brands of low-fat ice cream. Seventeen of them prefer Brand B, and 7 prefer Brand A. Using the sign test at the .05 significance level, can you conclude that among all people there is a preference for Brand B?

15.84 Four hundred randomly selected football fans were asked whether they prefer watching college football or professional football. Of these fans, 220 said they prefer the professional games, 168 prefer college games, and 12 have no preference. At the 2% level of significance, can you conclude that among all football fans there is a preference for either professional or college football?

15.85 A random sample of 200 customers of a large bank are asked whether they prefer using an automatic teller machine (ATM) or seeing a human teller for deposits and withdrawals. Of these customers, 122 said they prefer an ATM, 66 prefer a teller, and 12 have no opinion. At the 1% level of significance, can you conclude that more than half of all customers of this bank prefer an ATM?

15.86 Suppose that a polling agency is conducting a telephone survey. When prospective participants answer the phone, they are told that the survey will take just five minutes of their time. Ten randomly selected calls are monitored. The lengths of time (in minutes) required for the survey in these 10 cases are shown here.

7.1	6.3	4.9	5.0	5.7	9.0	8.2	5.9	6.5	7.7

Using the sign test at the 5% level of significance, can you conclude that the median time for the survey exceeds 5 minutes?

15.87 In 2001, the median age of buyers of Harley-Davidson motorcycles was 45 years (*USA TODAY*, June 7, 2002). Suppose that a random sample of 25 persons who bought Harley-Davidson motorcycles recently showed that 16 of them were over 45 years of age, 7 were under 45, and 2 were 45 years old. At the 5% level of significance, can you conclude that the current median age of Harley-Davidson buyers is over 45 years?

15.88 A state motor vehicle department requires auto owners to bring their autos to state emission centers periodically for testing. State officials claim that the median waiting time between the hours of 8 A.M. and 11 A.M. on weekdays at a particular site is 25 minutes. In a check of 30 randomly selected motorists during this time period at this site, 9 motorists waited for less than 25 minutes, 2 waited exactly 25 minutes, and 19 waited longer than 25 minutes.

a. Using the sign test at the 5% significance level, can you conclude that the median waiting time at this site during these hours exceeds 25 minutes?
b. Perform the test of part a at the 2.5% level of significance.
c. Comment on the results of parts a and b.

15.89 The following data give the amounts (in dollars) spent on textbooks by 35 college students during the 2005–2006 academic year.

475	418	680	610	655	488	710	375	250
695	420	610	380	98	530	415	757	357
409	611	455	618	395	612	468	610	780
450	880	490	490	626	850	688	588	

Using $\alpha = .05$, can you conclude that the median expenditure on textbooks by all such students in 2005–2006 was different from \$650?

15.90 Two brothers, Bob and Morris, who work the same hours at the same company in a large city, share an apartment on the outskirts of the city. When weather permits, Bob rides his bicycle to work, but Morris always drives. Although they always leave for work at exactly the same time each morning, Morris often arrives later than Bob because of the heavy traffic. Last year, on 21 randomly selected days of good weather, Bob arrived at work first 16 times, and Morris was first on 5 days. At the 5% level of significance, can you conclude that the median morning commuting time for Bob is less than that for Morris?

15.91 Refer to Exercise 15.34 and to Exercise 10.96 of Chapter 10, which concern Gamma Corporation's installation of governors on its salespersons' cars to regulate their speeds. The following table gives the gas mileage (in miles per gallon) for each of seven sales representatives' cars during the week before governors were installed and the gas mileage in the week after installation.

Salesperson	A	B	C	D	E	F	G
Before	25	21	27	23	19	18	20
After	26	24	26	25	24	22	23

a. Using the sign test at the 5% level of significance, can you conclude that the use of governors tends to increase the median gas mileage for Gamma Corporation's sales representatives' cars?

b. Compare your conclusion of part a with the result of the Wilcoxon signed-rank test that was performed in part a of Exercise 15.34 and with the result of the corresponding hypothesis test (using the t distribution) of Exercise 10.96.

c. If there is a difference in the three conclusions, how can you account for it?

15.92 A reporter for a travel magazine wanted to compare the effectiveness of two large travel agencies (X and Y) in finding the lowest airfares to given destinations. She randomly chose 32 destinations from the many offered by both agencies. She and her assistants requested the lowest available fare for each destination from each agency. For 18 of these destinations Agency X quoted a fare lower than that of Agency Y, for 8 destinations Agency Y found a lower fare, and in 6 cases the fares were the same. At the 2% level of significance, can you conclude that there is any difference in the median fares quoted by Agency X and Agency Y for all destinations they both offer?

15.93 Thirty-five patients with high blood pressures are given medication to lower their blood pressures. For all 35 patients, their blood pressures are measured before they begin the medication and again after they have finished taking medication for 30 days. For 25 patients the blood pressures were lower after finishing the medication, in 7 cases they were higher, and for 3 patients there was no change. Assume that these 35 patients make up a random sample of all people suffering from high blood pressures. At the 2.5% level of significance, can you conclude that the median blood pressure in all such patients is lower after the medication than before?

15.94 The following table shows the one-week sales of six salespersons before and after they attended a course on "how to be a successful salesperson."

Before	12	18	25	9	14	16
After	18	24	24	14	19	20

a. Using the Wilcoxon signed-rank test at the 5% significance level, can you conclude that the weekly sales for all salespersons tend to increase as a result of attending this course?

b. Perform the test of part a using the sign test at the 5% significance level.

c. Compare your conclusions from parts a and b.

15.95 An official at a figure skating competition thinks that two of the judges tend to score skaters differently. Shown next are the two judges' scores for eight skaters.

Skater	1	2	3	4	5	6	7	8
Judge A	5.8	5.7	5.6	5.9	5.8	5.9	5.8	5.6
Judge B	5.4	5.5	5.7	5.4	5.6	5.3	5.4	5.6

Using the Wilcoxon signed-rank test at the 5% level of significance, can you conclude that either judge tends to give higher median scores than the other?

15.96 Refer to Exercise 15.26. Consider the data given in that exercise on the cholesterol levels (in milligrams per hundred milliliters) for 30 randomly selected adults as determined by two laboratories, A and B.

a. Using the Wilcoxon signed-rank test at the 1% level of significance, can you conclude that the median cholesterol level for all such adults as determined by Lab A is higher than that determined by Lab B?

b. Compare your conclusion in part a to that of Exercise 15.26.

15.97 A consumer agency conducts a fuel economy test on two new subcompact cars, the Mouse (M) and the Road Runner (R). Each of 18 randomly selected drivers takes both cars on an 80-mile test run. For each driver, the gas mileage (in miles per gallon) is recorded for both cars; then the gas mileage for the R car is subtracted from the gas mileage for the M car. Thus, a minus difference indicates a higher gas mileage for the R car. One of the 18 drivers obtains exactly the same gas mileage for both cars. For the other drivers, the differences are ranked. The sum of the positive ranks is 31, and the sum of the absolute values of the negative ranks is 122. Can you conclude that the R car gets better gas mileage than the M car? Use the 2.5% level of significance.

15.98 Each of the two supermarkets, Al's and Bart's, in River City claims to offer lower-cost shopping. Fifty people who normally do the grocery shopping for their families are chosen at random. Each shopper makes up a list for a week's supply of groceries. Then these items are priced and the total cost is computed for each store. The paired differences are then calculated for each of the 50 shoppers, where a paired difference is defined as the cost of a cart of groceries at Al's minus the cost of the same cart of groceries at Bart's. These paired differences were positive for 21 shoppers and negative for 29 shoppers. The sum of ranks of the positive paired differences was 527 and the sum of the absolute values of the ranks of the negative paired differences was 748. Using the 1% level of significance, can you conclude that either store is less expensive than the other?

15.99 A consumer advocate is comparing the prices of eggs at supermarkets in the suburbs with the prices of eggs at supermarkets in the cities. The following data give the prices (in dollars) of a dozen large eggs in 13 supermarkets, 6 of which are in cities and 7 are in suburbs.

City	1.49	1.29	1.35	1.58	1.33	1.47	
Suburb	.99	1.09	1.39	1.28	1.16	1.44	1.05

Using the .05 level of significance and the Wilcoxon rank sum test, can you conclude that egg prices tend to be higher in the city?

15.100 Many VCR owners have difficulty learning to program the VCR to record TV programs. A consumer magazine tested two new VCRs, Brands X and Y, which are claimed to be user-friendly by their manufacturers. A random sample of 13 adults (6 for Brand X, 7 for Brand Y) are observed to see how quickly they can learn to program the VCRs properly. The following table gives the times (in minutes).

Brand X	32	36	28	43	98	39	
Brand Y	33	18	21	25	24	27	17

Using $\alpha = .05$ and the Wilcoxon rank sum test, can you conclude that learning times tend to be longer for Brand X?

15.101 A researcher obtains a random sample of 24 students taking elementary statistics at a large university and divides them randomly into two groups. Group A receives instruction to use Software A to do a statistics assignment, whereas Group B is taught to use Software B to do the same statistics assignment. The time (in minutes) taken by each student to complete this assignment is given in the table.

Group A	123	101	112	85	87	133	129	114	150	110	180	115
Group B	65	115	95	100	94	72	60	110	99	102	88	97

a. Using the 5% level of significance and the Wilcoxon rank sum test, can you conclude that the median time required for all students taking elementary statistics at this university to complete this assignment is longer for Software A than for Software B?

b. Would a paired-samples sign test be appropriate here? Why or why not?

15.102 Refer to Exercise 15.101. The scores on the homework assignment for the 24 students are given in the table.

Group A	48	38	45	31	42	25	40	43	50	30	33	46
Group B	37	21	40	27	49	44	36	41	20	39	18	40

Using the 10% significance level and the Wilcoxon rank sum test, can you conclude that there is a difference in the median scores for all students using Software A and all students using Software B?

15.103 Manufacturers of luxury cars are very much interested in knowing the age distribution of their customers because then they can change these models to attract younger buyers without losing the older customers who have traditionally favored such cars. According to data from CNW Marketing Research, the median ages of drivers (primary drivers using vehicles for personal use only) of Rolls-Royce, Mercedes, and Cadillac automobiles were 62.9, 58.7, and 53.4 years, respectively, at the time of the survey (*USA TODAY*, February 17, 2005). The following table gives the ages of seven randomly selected primary drivers of each of these three makes of cars.

Rolls-Royce	Mercedes	Cadillac
64	61	52
61	47	63
70	66	39
68	71	55
55	44	50
64	53	47
68	58	61

At the 5% level of significance, can you reject the null hypothesis that the median age of drivers for each of these three makes of cars is the same?

15.104 An academic employment service compared the starting salaries of May 2005 graduates in three major fields. Random samples were taken of 8 engineering majors, 10 business majors, and 7 mathematics majors. The starting salaries of all 25 graduates were determined and then ranked, yielding the following rank sums:

Engineering: 137 Business: 126 Mathematics: 62

At the 5% level of significance, can you reject the null hypothesis that the median starting salary is the same for May 2005 graduates in these three fields?

15.105 A sports magazine conducted a test of three brands (A, B, and C) of golf balls by having a professional golfer drive six balls of each brand. The lengths of the drives (in yards) from this test are listed in the table.

Brand A	Brand B	Brand C
275	245	267
266	256	283
301	261	259
281	270	250
288	259	263
277	262	256

At the 5% level of significance, can you reject the null hypothesis that the median distance of drives by this golfer is the same for all three brands of golf balls?

15.106 A students' group at a state university wanted to compare textbook costs for students majoring in economics, history, and psychology. The group obtained data from random samples of 10 economics majors, 9 history majors, and 11 psychology majors, all in the second semester of their junior year. The total textbook costs of the 30 students were recorded and ranked. The rank sums for economics and history majors were 134 and 157, respectively.

a. Find the rank sum for psychology majors. [*Hint:* The sum of n integers from 1 through n is given by $n(n + 1)/2$.]

b. At the 2.5% significance level, can you reject the null hypothesis that the median textbook costs are the same for students in all three majors who are in the second semester of the junior year?

15.107 The following table shows the average verbal SAT score and the percentage of high school graduates who took the SAT in 2002 for a random sample of 10 states.

State	Average Verbal SAT Score	Percentage of Graduates Taking SAT
Connecticut	509	83
Georgia	489	65
Illinois	578	11
Kentucky	550	12
Michigan	558	11
New Jersey	498	82
South Carolina	488	59
South Dakota	576	5
Vermont	512	69
Wisconsin	583	7

Source: The College Board, *The World Almanac and Book of Facts, 2003*.

a. For all 50 states, would you expect ρ_s to be positive, negative, or near zero? Why?
b. Calculate r_s for the sample of 10 states and indicate whether its value is consistent with your answer to part a.
c. Using the value of r_s calculated in part b, test H_0: $\rho_s = 0$ against H_1: $\rho_s \neq 0$ using the 5% level of significance.

15.108 The Spearman rho rank correlation coefficient may be used in cases where data for one or both variables are given in the form of ranks. Suppose that a film critic views 10 randomly chosen new movies and ranks them, with a rank of 10 being assigned to the film that he thinks will have the highest box office receipts, a rank of 9 to the next most profitable, and so forth. Three months after each film is released, its total box office receipts (in millions of dollars) are tabulated. The following table shows the ranking and receipts for each of these 10 films.

Rank	7	3	10	1	4	5	2	6	8	9
Receipts	40	5	66	2	3	10	28	15	30	17

a. Calculate r_s for the sample of 10 films.
b. Using the 5% level of significance, test H_0: $\rho_s = 0$ against H_1: $\rho_s > 0$.
c. Based on your conclusion in part b, is there sufficient evidence that this critic can predict the box office performance of a film?

15.109 Refer to Example 13–8, which contained data on monthly auto insurance premiums and years of driving experiences. Those data are reproduced here.

Driving Experience (years)	Monthly Auto Insurance Premium
5	\$64
2	87
12	50
9	71
15	44
6	56
25	42
16	60

The estimated regression line was found to be $\hat{y} = 76.6605 - 1.5476x$ in Example 13–8 and the simple linear correlation coefficient for the sample data was $-.77$. The true regression slope B was found to be significantly less than zero at the significance level of 5%.

a. Based on this information, if ρ_s is the Spearman rho rank correlation coefficient for the entire population from which this sample was taken, what would you expect from a test of H_0: $\rho_s = 0$ against H_1: $\rho_s < 0$ at the significance level of 5%?

b. Perform the hypothesis test mentioned in part a.

15.110 A researcher wonders if men still tend to stand aside and let women board elevators ahead of them. She observes 10 men and 10 women boarding the same elevator. The order in which they boarded is given here.

W W W M M W W M W W W M M M M W W M M M

Using the 5% level of significance, can you conclude that the order of boarding is nonrandom with respect to gender?

15.111 A machinist is making precision cutting tools. Because of the exacting specifications for these tools, about 20% of them fail to pass inspection and are judged defective. The shop supervisor feels that the machinist tends to produce defective tools in clusters, perhaps due to fatigue or distraction. If this is true, then a sequence of tools produced by this machinist will tend to have fewer runs of defective and good tools than expected by chance. The supervisor chooses a day at random and observes the following sequence of 18 tools, where G denotes a tool that passed inspection and D indicates a defective tool.

G G G G G D D G G G G D G G G D D G

Do you think that there is evidence of nonrandomness in this sequence? Use the 5% level of significance.

15.112 Some states require periodic testing of cars to monitor the emission of pollutants. A state official suspects that the inspection process at a particular station is faulty, that a car's test result may be affected by the tests of preceding cars. Analysis of the sequence of test results for a random sample chosen on a day yields the following information:

$$n_1 = \text{number of cars that passed the test} = 157$$

$$n_2 = \text{number of cars that failed the test} = 143$$

$$R = \text{number of runs} = 41$$

Using the 1% level of significance, can you conclude that the test results for this emissions test station are not random?

15.113 The following data give the sequence of wins and losses in 30 consecutive games for a baseball team during a season.

L W L W L W L L W W W L L W L L L L L W L L L L L W L W L L

Can you conclude that the wins are randomly distributed for this baseball team? Use the 2% significance level.

Advanced Exercises

15.114 A medical researcher wants to study the effects of a low-calorie diet on the longevity of laboratory mice. She randomly divides 20 mice into two groups. Group A gets a standard diet, while Group B receives a diet that contains all the necessary nutrients but provides only 70% as many calories as Group A's diet. The experiment is conducted for 36 months and the length of life (in days) of each mouse is recorded. The data obtained on the lives of these mice are shown in the following table. In these data, the asterisk (*) indicates that this mouse was still alive at the end of the 36-month experiment.

Group A	900	907	751	833	920	787	850	877	848	901
Group B	1037	905	1023	988	1078	1011	*	1063	898	1033

a. Using the 2.5% level of significance and the Wilcoxon rank sum test, does the low-calorie diet appear to lengthen the longevity of laboratory mice? Should the mouse that was still alive at the end of the experiment be eliminated from your analysis or is there a way to include it?

b. Would a Wilcoxon signed-rank test be appropriate in this example? Why or why not?

15.115 The editor of an automotive magazine has asked you to compare the median gas mileages for driving in the city for three models of compact cars. The editor has made available to you one car of each of the three models, three drivers, and a budget sufficient to buy gas and pay the drivers for approximately 500 miles of city driving for each car.

a. Explain how you would conduct an experiment and gather the data for a magazine article comparing gas mileages.

b. Suppose you wish to test the null hypothesis that the median gas mileages for driving in the city are the same for all three models of cars. Outline the procedure for using your data to conduct this test. Do not assume that the gas mileages for all cars of each model are normally distributed.

15.116 Refer to Exercise 10.96 in Chapter 10. Suppose Gamma Corporation decides to test the governors on seven cars. However, management is afraid that the speed limit imposed by the governors will reduce the number of contacts the salespersons can make each day. Thus, both the fuel consumption and the number of contacts made are recorded for each car/salesperson for each week of the testing period, both before and after the installation of governors.

	Number of Contacts		Fuel Consumption (mpg)	
Salesperson	**Before**	**After**	**Before**	**After**
A	50	49	25	26
B	63	60	21	24
C	42	47	27	26
D	55	51	23	25
E	44	50	19	24
F	65	60	18	22
G	66	58	20	23

Suppose that you are directed to prepare a brief report that includes statistical analysis and interpretation of the data. Management will use your report to help decide whether or not to install governors on all salespersons' cars. Use 5% significance levels for any hypothesis tests you perform to make suggestions. In contrast to Exercise 10.96, do not assume that the numbers of contacts, fuel consumption, or differences are normally distributed.

15.117 Suppose that you are a newspaper reporter and your editor has asked you to compare the hourly wages of carpenters, plumbers, electricians, and masons in your city. Because many of these workers are not union members, the wages may vary considerably among individuals in the same trade.

a. What data should you gather to make this statistical analysis and how would you collect them? What sample statistics would you present in your article and how would you calculate them? Assume that your newspaper is not intended for technical readers.

b. Suppose that you must submit your findings to a technical journal that does require statistical analysis of your data. If you want to determine whether or not the median hourly wages are the same for all four trades, briefly describe how you would analyze the data. Do not assume that the hourly wages for these populations are normally distributed.

15.118 Consider the data in the following table.

x	10	20	30	40	50	60
y	12	15	19	21	25	30

a. Suppose that each value of y in the table is increased by 5 but the x values remain unchanged. What effect will this have on the rank of each value of y? Do you expect the value of r_s to increase, decrease, or remain the same? Explain why.

b. Now, first calculate the value of r_s for the data in the table, and then increase each value of y by 5 and recalculate the value of r_s. Does the value of r_s increase, decrease, or remain the same? Does this result agree with your expectation in part a?

15.119 The English department at a college has hired a new instructor to teach the composition course to first-year students. The department head is concerned that the new instructor's grading practices might not be consistent with those of the professor (let us call him Professor A) who taught this course previously. She randomly selects 10 essays written by students for this class and makes two copies of each essay. She asks Professor A and this instructor (working independently) to assign a numerical grade to each of the 10 essays. The results are shown in the following table.

Essay	1	2	3	4	5	6	7	8	9	10
Professor A	75	62	90	48	67	82	94	76	78	84
Instructor	80	50	85	55	63	78	89	81	75	83

a. Suppose the department head wants to determine whether the instructor tends to grade higher or lower than Professor A. Which of the statistical tests discussed in this chapter could she use? Note that more than one test may be appropriate.

b. Using an appropriate test from your answer in part a, can you conclude that the instructor tends to grade higher or lower than Professor A? Use $\alpha = .05$.

c. Suppose the department head wants to determine whether the instructor is consistent with Professor A in the sense that they tend to agree on which paper is best, which is second best, and so forth. Which test from this chapter would be appropriate to use? State the relevant null and alternative hypotheses.

d. Using the test you chose in part c, can you conclude that Professor A and the instructor are consistent in their grading? Use the 5% level of significance.

15.120 Three doctors are employed at a large clinic. The manager at the clinic wants to know whether these three doctors spend the same amount of time per patient. The manager randomly chooses 10 routine appointments of patients with each of the three doctors and times them. Thus, the data set consists of 10 observations on the time spent with patients by each doctor.

a. To test the null hypothesis that the mean or median times are equal for all three doctors against the alternative hypothesis that they are not all equal, which tests from Chapters 12 and 15 are appropriate?

b. For each test that you indicated in part a, specify whether the test is about the means or the medians.

c. What assumptions are required for the test from Chapter 12?

15.121 An educational researcher is studying the relationship between high school grade point averages (GPAs) and SAT scores. She obtains GPAs and SAT scores for a random sample of 25 students and wants to test the null hypothesis that there is no correlation between GPAs and SAT scores against the alternative hypothesis that these variables are positively correlated.

a. If she wants to base her test on the linear correlation coefficient of Chapter 13, what assumptions are required about the two variables (GPAs and SAT scores)?

b. If the assumptions required in part a are not satisfied, what other test(s) might she use?

15.122 To test the effectiveness of a new six-week body-building course, 12 tenth-grade boys are randomly selected. Each boy is tested before and after the course to see how much weight he can lift.

a. To test whether or not the mean or median weight lifted by all such boys tends to be greater after the course than before, which tests from Chapters 10 and 15 might be used?

b. For each test that you indicated in part a, specify whether it involves the mean or median.

c. If the paired differences in weights lifted before and after the test are not normally distributed, which of the tests indicated in part a could be used?

15.123 Suppose in a sample we have 10 A's and 15 B's. What is the maximum number of runs possible in a sequence of these 25 letters?

15.124 Refer to Exercises 12.27 and 15.56. In these two problems you were asked to perform an ANOVA and a Kruskal-Wallis test, respectively, on the data. In both cases, the results were significant at the 5% significance level. Change the values of the tips in those data so that the p-value for the Kruskal-Wallis test remains the same, but the ANOVA results are no longer significant at the 5% level. (Hint: When making the changes, the ranks of the fifteen data points should not change.)

15.125 A student who typically does not do his homework was asked to toss a coin 20 times and write down the sequence of results. Instead of tossing the coin, the student simply wrote down the following sequence (reading from left to right) of hypothetical outcomes.

H	T	H	T	H	T	H	H	T	H
T	T	H	T	H	T	H	T	H	T

Use the appropriate test to show that the professor was justified in accusing the student of not actually tossing the coin.

Self-Review Test

1. Nonparametric tests
- **a.** are more efficient than the corresponding parametric tests
- **b.** do not require that the population being sampled has a normal distribution
- **c.** generally require more assumptions about the population than parametric tests do

2. For small samples ($n \leq 25$), the critical value(s) for the sign test are based on the
a. binomial distribution **b.** normal distribution **c.** t distribution

3. Which of the following tests may be used to test hypotheses about one median?
a. The sign test **b.** The Kruskal-Wallis test **c.** The Wilcoxon rank sum test

4. The Wilcoxon signed-rank test may be used to test
- **a.** for a difference between the medians of two independent samples
- **b.** for a preference for one product over another
- **c.** hypotheses involving paired samples

5. When we use the Wilcoxon signed-rank test,
- **a.** all observations are ranked
- **b.** the difference for each pair is calculated and then all the differences are ranked according to their absolute values
- **c.** only the signs of the differences are used to calculate the value of the test statistic

6. In order to perform a Wilcoxon rank sum test, one must calculate the
- **a.** standard deviation of each sample
- **b.** range of the data
- **c.** rank of each observation

7. Which of the following tests may be used with paired samples? Circle all that apply.
- **a.** Sign test
- **b.** Wilcoxon signed-rank test
- **c.** Wilcoxon rank sum test
- **d.** Spearman rho rank correlation coefficient test

8. The Spearman rho rank correlation coefficient is calculated as the
- **a.** simple linear correlation coefficient between the two sets of observations
- **b.** simple linear correlation coefficient between the ranks of the two sets of observations
- **c.** square of the simple linear correlation coefficient between two sets of observations

9. In order to test a hypothesis about the Spearman rho rank correlation coefficient
- **a.** both sets of data must come from normally distributed populations
- **b.** one set of data must come from a normally distributed population
- **c.** either set of data can have any distribution

10. The Spearman rho rank correlation coefficient is positive when
- **a.** there is no relationship between the two sets of observations
- **b.** the values in one set of observations increase as the values of the corresponding observations in the other set decrease
- **c.** the values in one set of observations increase as the values of the corresponding observations in the other set increase

11. For the runs test for randomness, which of the following statements are true?
- **a.** We notice which of the two possible outcomes has occurred at each stage in a list of consecutive outcomes.
- **b.** A run is one or more consecutive occurrences of either one of the two possible outcomes.
- **c.** We are testing the hypothesis that one of the two possible outcomes occurred significantly more frequently than the other.

12. In the runs test for randomness, we reject the null hypothesis
- **a.** only when there is a very large number of runs
- **b.** only when there is a very small number of runs
- **c.** if there is either a very large or a very small number of runs

13. In the runs test for randomness, the distribution of R (the total number of runs) is approximately normal when
- **a.** R is greater than 10
- **b.** at least one of the two possible outcomes occurs more than 15 times
- **c.** each of the two possible outcomes occurs more than 15 times

14. A large pool of prospective jurors is made up of an equal number of men and women. A 12-person jury selected from this pool consists of 2 women and 10 men. At the 5% level of significance, can we reject the null hypothesis that the selection process is unbiased in terms of gender?

15. A September 2002 *USA TODAY*/Gallup poll asked Americans whether they favored a proposal to put Social Security payroll taxes into personal retirement accounts. Fifty-two percent of the respondents were in favor of the proposal (*USA TODAY*, September 25, 2002). Suppose that the 2002 poll consisted of 1000 respondents, so that 520 favored the proposal. Using the 2.5% level of significance, can you conclude that more than half of all Americans favor putting Social Security payroll taxes into personal retirement accounts?

16. The past records of a supermarket show that its customers spent a median of $65 per visit. After a promotional campaign designed to increase spending, the store took a sample of 12 customers and recorded the amounts (in dollars) they spent. The amounts are listed here.

88 69 141 28 106 45 32 51 78 54 110 83

Using $\alpha = .05$, can you conclude that the median amount spent by all customers at this store after the campaign exceeds $65?

17. According to a U.S. Census Bureau survey of households, women living alone had a median income of $20,264 in 2001 (*USA TODAY*, September 25, 2002). Suppose that in a recent random sample of 400 women living alone, 229 had incomes under $20,264 and 171 had incomes over $20,264. At the 1% level of significance, can you conclude that the median income of women living alone currently is different from $20,264?

18. The following table gives the cholesterol levels for seven adults before and after they completed a special dietary plan.

Before	210	180	195	220	231	199	224
After	193	186	186	223	220	183	233

a. Using the sign test at the 5% significance level, can you conclude that the median cholesterol levels are the same before and after the diet?
b. Using the Wilcoxon signed-rank test at the 5% significance level, can you conclude that the median cholesterol levels are the same before and after the diet?
c. Compare your conclusions for parts a and b.

19. An archeologist wants to compare two methods (I and II) of radioactive dating of artifacts. He submits a random sample of 33 artifacts that are suitable for radioactive dating. Each one of these artifacts is dated by both methods. The paired differences are then calculated for each of the 33 artifacts, where a paired difference is defined as the age of an artifact dated by Method I minus the age of the same artifact dated by Method II. These paired differences were positive for 11 of the artifacts, negative for 20, and zero for 2 artifacts. Using the sign test at the 2% level of significance, can you conclude that the median estimated ages of such artifacts differ for the two methods?

20. A professor at a large university suspects that the grades of engineering majors tend to be lower in the spring semester than in the fall semester. He randomly selects 10 sophomore electrical engineering majors and records their grade point averages (GPAs) for the fall and the spring semesters. The data obtained are shown in the table.

Student	1	2	3	4	5	6	7	8	9	10
Fall GPA	3.20	3.56	3.05	3.78	4.00	2.85	3.33	2.67	3.00	3.67
Spring GPA	3.15	3.40	2.88	3.67	4.00	3.00	3.30	3.05	2.95	3.50

Using the Wilcoxon signed-rank test at the 5% level of significance, can you conclude that the median GPA of all sophomore electrical engineering majors at this university tends to be lower in the spring semester than in the fall semester?

21. A random sample of 30 students was selected to test the effectiveness of a course designed to improve memory. Each student was given a memory test before and after taking the course. Each student's score after taking the course was subtracted from his or her score before the course; then the 30 differences were ranked. Thus, a negative rank denotes an improved score after taking the course. The sum of the positive ranks was 102; the sum of the absolute values of the negative ranks was 276. Three students scored exactly the same on both tests. Using the 2.5% level of significance, can you conclude that the course tends to improve scores on memory tests?

22. A commuter has two alternative routes (Route 1 and Route 2) to drive to work. Picking days at random, she drives to work using each route for eight days and records the time (in minutes) taken to commute from home to work on each day. These times are shown in the following table.

Route 1	45	43	38	56	41	43	46	44
Route 2	38	40	39	42	50	37	46	36

Using the Wilcoxon rank sum test at the 5% level of significance, can you reject the null hypothesis that the median commuting time is the same for both routes?

23. An accounting firm has hired two temporary employees, A and B, to prepare individual federal income tax returns during the tax season. Clients who have relatively simple tax situations are randomly assigned to either A or B. The firm randomly selected 18 income tax returns prepared by each of these two employees and recorded the times taken to prepare these tax returns. After these times taken to prepare 36 tax returns were ranked, the sum of the ranks for A was found to be 298 and the sum of the ranks for B was equal to 368. Using the Wilcoxon rank sum test at the 2.5% level of significance, can you conclude that there is a difference in the median times taken to prepare such income tax returns by A and B?

24. The following table lists the numbers of cases of telemarketing fraud reported to law-enforcement officials during several randomly chosen weeks in 2002 for three large cities of approximately equal populations.

City A	City B	City C
53	29	75
46	35	49
59	44	62
33	31	68
60	50	52
	48	

a. At the 2.5% level of significance, can you reject the null hypothesis that the distributions of the numbers of such reported cases are identical for all three cities?
b. Can you reject the null hypothesis of part a at the 1% level of significance?
c. Comment on the results of parts a and b.

25. The following is a list of home runs (denoted by x) and runs batted in (denoted by y) as of July 1, 2005, by 10 players selected at random from a minor league baseball team.

Player	1	2	3	4	5	6	7	8	9	10
x	10	7	13	2	8	4	16	11	5	4
y	49	38	54	20	41	27	62	40	22	19

a. As home runs increase, runs batted in tend to increase. From this, do you expect the value of the Spearman rho rank correlation coefficient to be positive or negative?
b. Compute r_s for the data.
c. Suppose ρ_s is the value of the Spearman rho rank correlation coefficient for all players in this league. Using the 2.5% significance level, test the null hypothesis H_0: $\rho_s = 0$ against the alternative hypothesis H_1: $\rho_s > 0$.

26. Ramon fishes in a lake where the minimum size for bass to be kept is 12 inches long; all smaller bass must be returned to the water. He thinks that most of the "keepers" (bass 12 inches or longer) are caught early in the morning. If he is right, there should be a few long runs of keepers caught in the early morning followed by a few long runs of smaller bass caught later on during the day. Thus, if the fish are recorded in sequence, there should be fewer runs than expected by chance. Last Saturday Ramon fished from 6 A.M. to 11 A.M. and caught 14 bass in the following order, where K denotes a keeper and S denotes a bass shorter than 12 inches.

K K K K S K K S K S S S S S

Using the 5% significance level, does this sequence support Ramon's theory?

27. As of June 1, 2005, a minor league baseball team had played 54 games, winning 30 and losing 24. In these 54 consecutive games, there were 15 runs (in the statistical sense). Using the runs test for randomness, can we conclude that the 30 wins and 24 losses are randomly spread out among the 54 games? Use a significance level of 5%.

Mini-Projects

MINI-PROJECT 15–1

For a period of 30 business days, record the daily price of crude oil and the price of a stock that you think might be affected by the oil price (for example, an oil company or an alternative energy company).

a. Compute the Spearman rho rank correlation coefficient for these data.
b. Can you conclude that there is a relationship between the two sets of prices? Use $\alpha = .05$.

MINI-PROJECT 15–2

For a period of 30 business days, record whether the Dow Jones Industrial Average moves up or down. Use your data to perform an appropriate test at the 1% level of significance to see if the sequence of upward and downward movements of the Dow appears to be random over this period. (As an alternative, you might use the NASDAQ index, the price of an individual stock, or the price of gold.)

MINI-PROJECT 15–3

On December 8, 2005, the U.S. national debt was approximately $8.13 trillion. Take random samples of 10 or more students from each of three different majors and ask each student to estimate the size of the national debt.

a. At the 5% level of significance and using the Kruskal-Wallis test, can you conclude that the median perceived values of the national debt are the same for all three majors?
b. Find the current value of the national debt. On the whole, did the students in your sample tend to overestimate its value, or did they tend to underestimate?

DECIDE FOR YOURSELF

Using Nonparametric Methods

Prior to this chapter, you used inferential methods to work with quantitative data that are classified as scale data or with categorical data that are classified as nominal data.[2] A third type of data, called ordinal data, often requires nonparametric methods for analysis. Ordinal data are data that can be ranked. For example, insurance companies will classify policy holders by their age groups, as opposed to their specific ages, for assessing risk. Since nobody can be in two different age groups at the same time, the data identifying age groups can be ranked.

Suppose you are interested in studying the relationship between the successes of Division I men's and women's basketball teams at colleges that have both of these sports. We can use Spearman's Rho to calculate the rank correlation between the power index rankings (RPIs) of the men's and women's teams. We have provided three scatterplots here. Specifically we have the scatterplot of the ranks of the men's teams versus the ranks of the women's teams for various colleges (Figure 15.15), the scatterplot of the RPIs of men's teams versus the RPIs of women's teams (Figure 15.16), and the scatterplot of the RPIs of men's teams versus their ranks (Figure 15.17). These data are for all Division I programs that had men's and women's basketball programs during the 2004–2005 season.

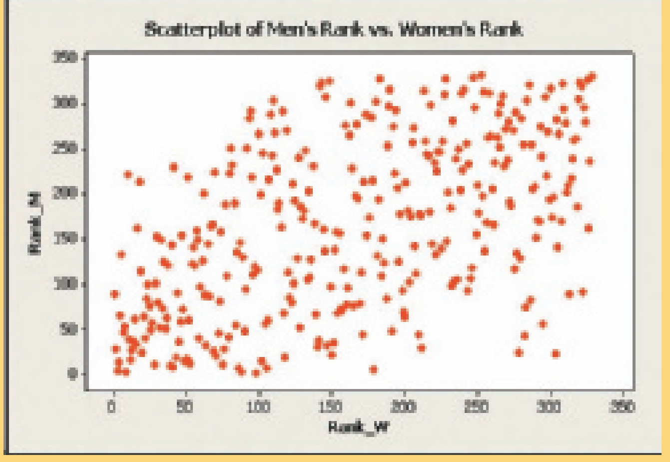


Figure 15.15

Answer the following questions.

1. Which of the three relationships should have its strength measured by Spearman's Rho rank correlation instead of the Pearson correlation coefficient?

[2]Based on what are called the scales or levels of measurement, data can be classified into four scales or levels—nominal, ordinal, interval, and ratio scales. Data that can be divided into different categories only for identification purposes are said to have a nominal scale. An example of such data is the names given to different makes of cars, such as Town Car, Toyota Camry, and so forth. Data that can be divided into different categories so that categories can be ranked are said to have an ordinal scale. An example of such data is an evaluation of a product as excellent, good, or poor. Data that can be ranked and for which the difference between any two values can be calculated (and is meaningful) are said to have an interval scale. An example of such data is temperatures in two cities. Data that can be ranked and for which all arithmetic operations (such as addition, subtraction, multiplication and division) can be done are said to have ratio scale. An example of such data is gross sales of two companies.

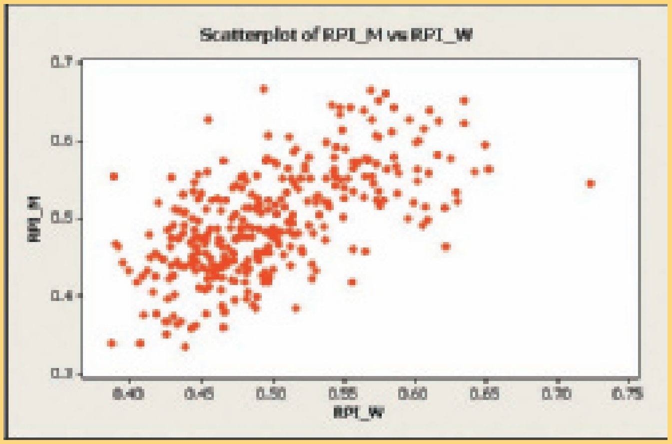


Figure 15.16

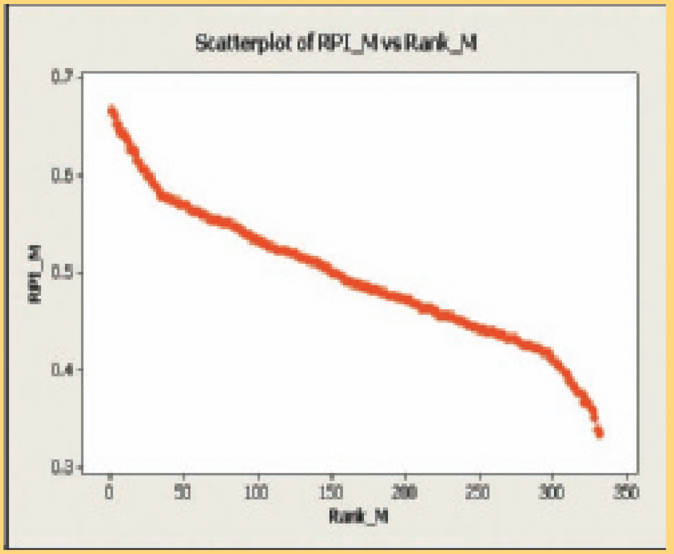


Figure 15.17

2. Which of the three relationships is inappropriate to be discussed in terms of its correlation?

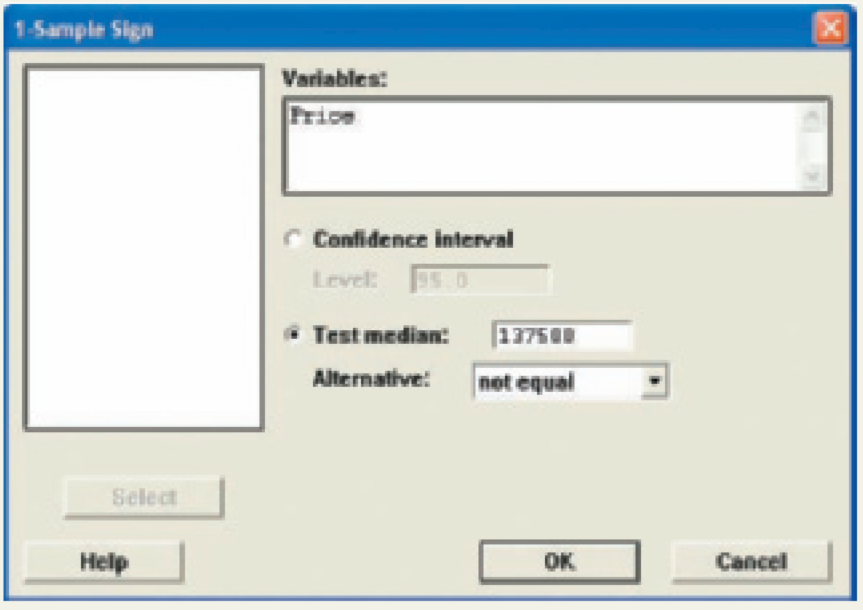
3. There is one point that stands out on the scatterplot of RPI values (Figure 15.16). Identify the location of this point on the scatterplot of ranks (Figure 15.15).

TECHNOLOGY INSTRUCTION Nonparametric Methods

TI-84

1. The TI-84 does not contain any built-in nonparametric methods.

MINITAB



Screen 15.1

1. To perform a sign test about a population median, select **Stat>Nonparametrics>1-Sample Sign**. Enter the name of the column containing your sample data in the box below **Variables**, select **Test median**, enter the hypothesized value of the median, and select your alternative hypothesis. Click **OK** to see the results. (See Screens 15.1 and 15.2.) Then use the *p*-value obtained in this output to make a decision.

2. To run a Wilcoxon rank sum test to determine if the populations from which two independent samples are drawn are identical, select **Stat>Nonparametrics>Mann-Whitney**. (Note that MINITAB does not have procedures for the Wilcoxon rank sum test. The Mann-Whitney test is very

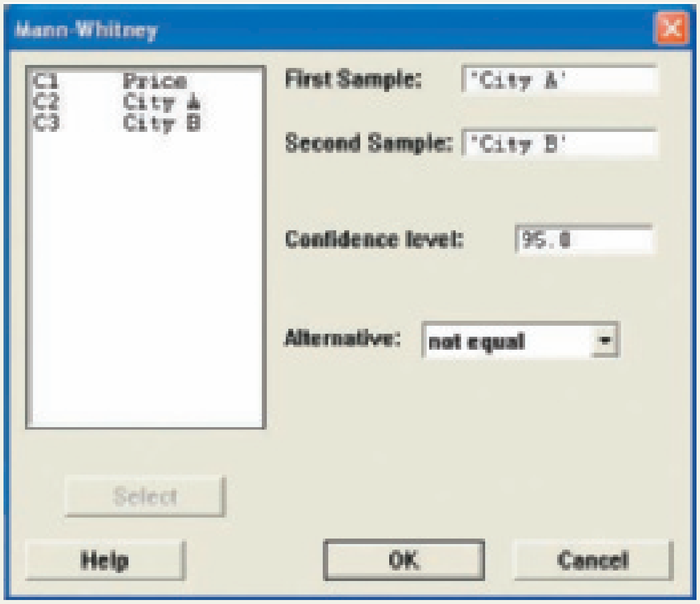
Sign Test for Median: Price

```
Sign test of median =  137500 versus not = 137500

        N  Below  Equal  Above       P  Median
Price  10      3      0      7  0.3438  151950
```

Screen 15.2

similar to the Wilcoxon rank sum test and we can use it here.) Enter the names of the two columns containing your data in their respective boxes. Select the alternative hypothesis. Click **OK** to see the results. (See Screens 15.3 and 15.4.) Then use the *p*-value obtained in this output to make a decision.

3. To run a Kruskal-Wallis test to determine if three or more populations have identical distributions, enter the data on the response variable in one column and the factor in another column. Select **Stat>Nonparametrics>Kruskal-Wallis** and enter the columns containing the response variable and factor in their respective boxes. Click **OK** to see the results. (See Screens 15.5 and 15.6.) Then use the *p*-value obtained in this output to make a decision.
4. To perform a Wilcoxon signed-rank test, enter the pairwise differences in a single column, making sure that you calculate each difference in the same order. Select **Stat>Nonparametrics>1-Sample Wilcoxon**. Enter the name of the column containing the differences in the box below **Variables**, select **Test median**, enter the hypothesized value of the median, and select your alternative hypothesis. Click **OK** to

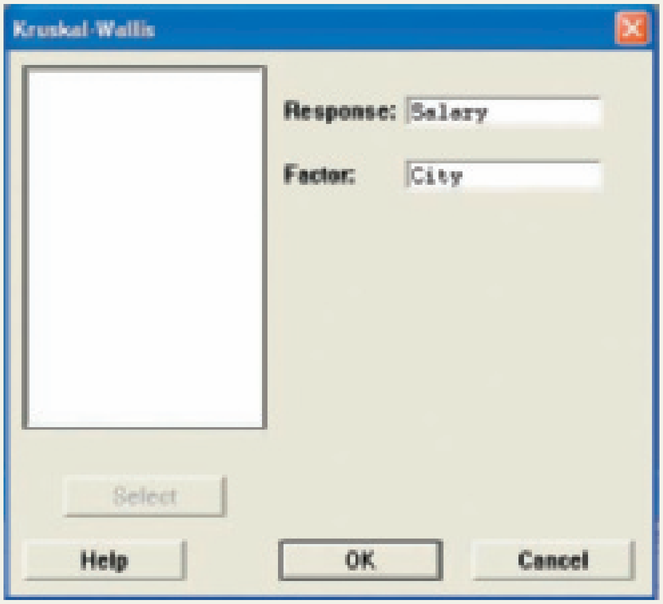


Screen 15.3

Mann-Whitney Test and CI: City A, City B

```
        N  Median
City A  8   17.50
City B  9   20.00

Point estimate for ETA1-ETA2 is -4.00
95.1 Percent CI for ETA1-ETA2 is (-11.00,3.00)
W = 58.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.2110
The test is significant at 0.2101 (adjusted for ties)
```

Screen 15.4



Screen 15.5

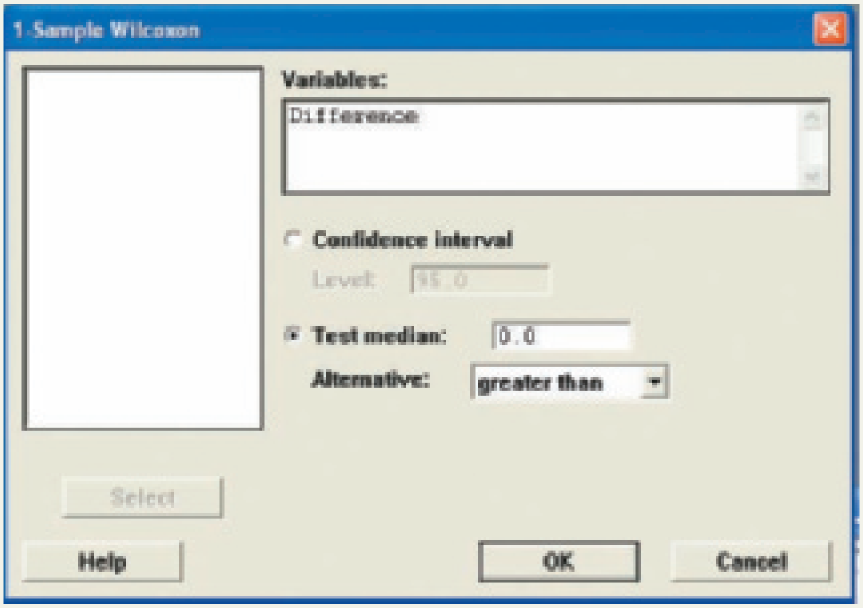
Kruskal-Wallis Test: Salary versus City

```
Kruskal-Wallis Test on Salary

City            N  Median  Ave Rank      Z
Atlanta         8   53.00      11.9   0.54
Boston          6   48.50      12.5   0.78
San Francisco   7   43.00       8.6  -1.23
Overall        21              11.0

H = 1.54  DF = 2  P = 0.462
H = 1.55  DF = 2  P = 0.462  (adjusted for ties)
```

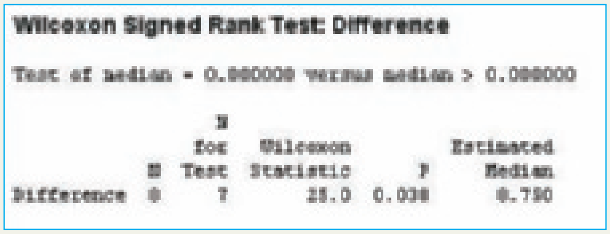
Screen 15.6

see the results. (See Screens 15.7 and 15.8.) Then use the *p*-value obtained in this output to make a decision.

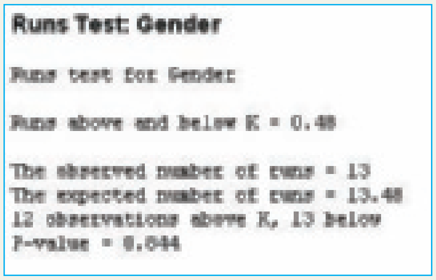
5. To perform a runs test to see if a set of data is random, enter the given data into a column. Note that categorical data must be given numeric values. Select **Stat>Nonparametrics>Runs Test** and enter the column containing data in the box below **Variables**. Click **OK** to see the results. (See Screens 15.9 and 15.10.) Then use the *p*-value obtained in this output to make a decision.



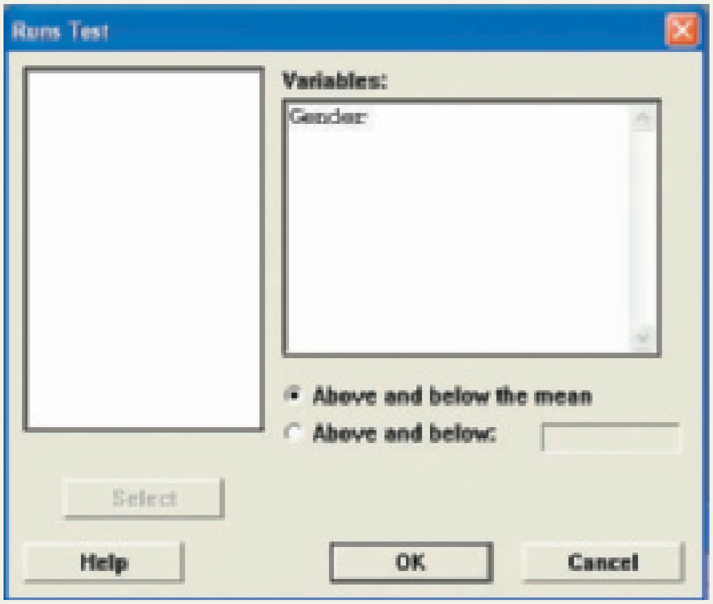
Screen 15.7



Screen 15.8



Screen 15.10



Screen 15.9

Excel

Excel does not contain any built-in nonparametric methods.

TECHNOLOGY ASSIGNMENTS

TA15.1 Fifteen coffee drinkers are selected at random and asked to test and state their preferences for Brand X coffee, Brand Y coffee, or neither (N). The results are as follows:

X X Y X N Y Y X Y X X Y Y X X

Let p be the proportion of coffee drinkers in the population who prefer Brand X. Using the sign test, perform the test H_0: $p = .50$ against H_1: $p > .50$. Use a significance level of 2.5%.

TA15.2 Twelve sixth-graders were selected at random and asked how many hours per week they spend watching television. The data obtained are shown here.

23	30	22.5	28	29	24.5	25	32	31	26	27	21

Using the sign test, can you conclude that the median number of hours spent per week watching television by all sixth-graders is less than 28? Use a significance level of 5%.

TA15.3 The manufacturer of an engine oil additive, Hyper-Slick, claims that this product reduces the engine friction and, consequently, increases the miles per gallon (mpg). To test this claim, 10 cars are driven on a fixed 300-mile course without the oil additive, and each car's mpg is calculated and recorded. Then the engine oil additive is added to each car and the process is repeated. The data obtained are shown in the table.

Car	MPG without Additive	MPG with Additive
1	20.00	19.90
2	23.60	27.85
3	29.40	28.70
4	25.70	28.20
5	35.80	37.30
6	32.20	31.30
7	26.30	26.10
8	31.80	36.80
9	29.00	32.75
10	24.70	29.20

Using the sign test, can you conclude that the manufacturer's claim is true? Use a significance level of 5%.

TA15.4 Do Technology Assignment TA15.3 using the Wilcoxon signed-rank test and a significance level of 5%. Compare your conclusion with that of Technology Assignment TA15.3 and comment.

TA15.5 Refer to Exercise 15.43. In a winter Olympics trial for women's speed skating, seven skaters use a new type of skate, while eight others use the traditional type. Each skater is timed (in seconds) in the 500-meter event. The results are given in the table.

New skates	40.5	40.3	39.5	39.7	40.0	39.9	41.5	
Traditional skates	41.0	40.8	40.9	39.8	40.6	40.7	41.1	40.5

Assume that these 15 skaters make up a random sample of all Olympic-class 500-meter female speed skaters. Using the Wilcoxon rank sum test (Mann-Whitney test), can you conclude that the new skates tend to produce faster times in this event? Use the 5% significance level.

TA15.6 Refer to Exercise 15.46. Two brands of tires are tested to compare their durability. Eleven Brand X tires and 12 Brand Y tires are tested on a machine that simulates road conditions. The mileages (in thousands of miles for each tire) are shown in the following table.

Brand X	51	55	53	49	50.5	57	54.5	48.5	51.5	52	53.5	
Brand Y	48	47	54	55.5	50	51	46	49.5	52.5	51	49	45

Using the Wilcoxon rank sum (Mann-Whitney) test, can you conclude that the median mileage for Brand X tires is greater than the median mileage for Brand Y tires? Use the 5% level of significance.

TA15.7 Three brands of 60-watt lightbulbs—Brand A, Brand B, and a generic brand—are tested for their lives. The following table shows the lives (in hours) of these bulbs.

Brand A	Brand B	Generic
975	1001	899
1050	1099	789
890	915	824
933	959	1011

962	986	907
925	957	923
1007	987	937
855	881	865
	1025	1024

Using the Kruskal-Wallis test with a significance level of .05, can you conclude that the distributions of the lives of lightbulbs are the same for all three brands?

TA15.8 Refer to Exercise 15.54. A consumer agency investigated the premiums charged by four auto insurance companies. The agency randomly selected five drivers insured by each company who had similar driving records, autos, and insurance coverages. The following table gives the monthly premiums paid by these 20 drivers.

Company A	Company B	Company C	Company D
\$65	\$48	\$57	\$62
73	69	61	53
54	88	89	45
43	75	77	51
70	72	69	44

Using the Kruskal-Wallis test at the 5% significance level, can you reject the null hypothesis that the distributions of auto insurance premiums paid per month by all such drivers are the same for all four companies?

TA15.9 Refer to Exercise 15.75. A fair coin is tossed 20 times in the presence of a psychic who claims that she can cause a nonrandom sequence of heads (H) and tails (T) to appear. The following sequence of heads and tails is obtained in these 20 tosses.

H H T H T H T T H T H H T T H H H T T H

Using the runs test, can you conclude that the psychic's claim is true? Use a significance level of 5%.

TA15.10 At a small soda factory, the amount of soda put into each 12-ounce bottle by the bottling machine varies slightly for each filling. The plant manager suspects that the machine has a random pattern of overfilling and underfilling the bottles. The following are the results of filling 18 bottles, where O denotes 12 ounces or more of soda in a bottle and U denotes less than 12 ounces of soda.

U U U O O O O U U O O O U O U U U U

Using the runs test at the 5% significance level, can you conclude that there is a nonrandom pattern of overfilling and underfilling such bottles?

Table VIII Critical Values of *X* for the Sign Test

	One tail $\alpha = .005$ Two tail $\alpha = .01$		One tail $\alpha = .01$ Two tail $\alpha = .02$		One tail $\alpha = .025$ Two tail $\alpha = .05$		One tail $\alpha = .05$ Two tail $\alpha = .10$	
n	Lower critical value	Upper critical value	Lower critical value	Upper critical value	Lower critical value	Upper critical value	Lower critical value	Upper critical value
1	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—
3	—	—	—	—	—	—	—	—
4	—	—	—	—	—	—	—	—
5	—	—	—	—	—	—	0	5
6	—	—	—	—	0	6	0	6
7	—	—	0	7	0	7	0	7
8	0	8	0	8	0	8	1	7
9	0	9	0	9	1	8	1	8
10	0	10	0	10	1	9	1	9
11	0	11	1	10	1	10	2	9
12	1	11	1	11	2	10	2	10
13	1	12	1	12	2	11	3	10
14	1	13	2	12	2	12	3	11
15	2	13	2	13	3	12	3	12
16	2	14	2	14	3	13	4	12
17	2	15	3	14	4	13	4	13
18	3	15	3	15	4	14	5	13
19	3	16	4	15	4	15	5	14
20	3	17	4	16	5	15	5	15
21	4	17	4	17	5	16	6	15
22	4	18	5	17	5	17	6	16
23	4	19	5	18	6	17	7	16
24	5	19	5	19	6	18	7	17
25	5	20	6	19	7	18	7	18

Source: D. B. Owen, *Handbook of Statistical Tables*. © 1962 by Addison-Wesley Publishing Company, Inc. Reprinted by permission of Addison Wesley Longman.

Table IX Critical Values of T for the Wilcoxon Signed-Rank Test

n	One-tailed $\alpha = .005$ Two-tailed $\alpha = .01$	One-tailed $\alpha = .01$ Two-tailed $\alpha = .02$	One-tailed $\alpha = .025$ Two-tailed $\alpha = .05$	One-tailed $\alpha = .05$ Two-tailed $\alpha = .10$
1	—	—	—	—
2	—	—	—	—
3	—	—	—	—
4	—	—	—	—
5	—	—	—	1
6	—	—	1	2
7	—	0	2	4
8	0	2	4	6
9	2	3	6	8
10	3	5	8	11
11	5	7	11	14
12	7	10	14	17
13	10	13	17	21
14	13	16	21	26
15	16	20	25	30

Source: Some Rapid Approximate Statistical Procedures, 1964. Reprinted with permission of Lederle Pharmaceutical Division of American Cyanamid Company, Philadelphia, PA.

Table X Critical Values of T for the Wilcoxon Rank Sum Test

a. One-tailed $\alpha = .025$; Two-tailed $\alpha = .05$

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

b. One-tailed $\alpha = .05$; Two-tailed $\alpha = .10$

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Source: Some Rapid Approximate Statistical Procedures, 1964. Reprinted with the permission of Lederle Pharmaceutical Division of American Cyanamid Company, Philadelphia, PA.

Table XI Critical Values for the Spearman Rho Rank Correlation Coefficient Test

	One-tailed α			
	.05	.025	.01	.005
	Two-tailed α			
n	.10	.05	.02	.01
5	$\pm$.900	—	—	—
6	$\pm$.829	$\pm$.886	$\pm$.943	—
7	$\pm$.714	$\pm$.786	$\pm$.893	$\pm$.929
8	$\pm$.643	$\pm$.738	$\pm$.833	$\pm$.881
9	$\pm$.600	$\pm$.700	$\pm$.783	$\pm$.833
10	$\pm$.564	$\pm$.648	$\pm$.745	$\pm$.794
11	$\pm$.536	$\pm$.618	$\pm$.709	$\pm$.755
12	$\pm$.503	$\pm$.587	$\pm$.678	$\pm$.727
13	$\pm$.475	$\pm$.566	$\pm$.672	$\pm$.744
14	$\pm$.456	$\pm$.544	$\pm$.645	$\pm$.714
15	$\pm$.440	$\pm$.524	$\pm$.622	$\pm$.688
16	$\pm$.425	$\pm$.506	$\pm$.601	$\pm$.665
17	$\pm$.411	$\pm$.490	$\pm$.582	$\pm$.644
18	$\pm$.399	$\pm$.475	$\pm$.564	$\pm$.625
19	$\pm$.388	$\pm$.462	$\pm$.548	$\pm$.607
20	$\pm$.377	$\pm$.450	$\pm$.534	$\pm$.591
21	$\pm$.368	$\pm$.438	$\pm$.520	$\pm$.576
22	$\pm$.359	$\pm$.428	$\pm$.508	$\pm$.562
23	$\pm$.351	$\pm$.418	$\pm$.496	$\pm$.549
24	$\pm$.343	$\pm$.409	$\pm$.485	$\pm$.537
25	$\pm$.336	$\pm$.400	$\pm$.475	$\pm$.526
26	$\pm$.329	$\pm$.392	$\pm$.465	$\pm$.515
27	$\pm$.323	$\pm$.384	$\pm$.456	$\pm$.505
28	$\pm$.317	$\pm$.377	$\pm$.448	$\pm$.496
29	$\pm$.311	$\pm$.370	$\pm$.440	$\pm$.487
30	$\pm$.305	$\pm$.364	$\pm$.432	$\pm$.478

Table XII Critical Values for a Two-Tailed Runs Test with $\alpha = .05$

n_1 \ n_2	5	6	7	8	9	10	11	12	13	14	15
2	—	—	—	—	—	—	—	2	2	2	2
								6	6	6	6
3	—	2	2	2	2	2	2	2	2	2	3
		8	8	8	8	8	8	8	8	8	8
4	2	2	2	3	3	3	3	3	3	3	3
	9	9	10	10	10	10	10	10	10	10	10
5	2	3	3	3	3	3	4	4	4	4	4
	10	10	11	11	12	12	12	12	12	12	12
6	3	3	3	3	4	4	4	4	5	5	5
	10	11	12	12	13	13	13	13	14	14	14
7	3	3	3	4	4	5	5	5	5	5	6
	11	12	13	13	14	14	14	14	15	15	15
8	3	3	4	4	5	5	5	6	6	6	6
	11	12	13	14	14	15	15	16	16	16	16
9	3	4	4	5	5	5	6	6	6	7	7
	12	13	14	14	15	16	16	16	17	17	18
10	3	4	5	5	5	6	6	7	7	7	7
	12	13	14	15	16	16	17	17	18	18	18
11	4	4	5	5	6	6	7	7	7	8	8
	12	13	14	15	16	17	17	18	19	19	19
12	4	4	5	6	6	7	7	7	8	8	8
	12	13	14	16	16	17	18	19	19	20	20
13	4	5	5	6	6	7	7	8	8	9	9
	12	14	15	16	17	18	19	19	20	20	21
14	4	5	5	6	7	7	8	8	9	9	9
	12	14	15	16	17	18	19	20	20	21	22
15	4	5	6	6	7	7	8	8	9	9	10
	12	14	15	16	18	18	19	20	21	22	22

Source: Frieda S. Swed and C. Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," *The Annals of Statistics* 14(1943). Reprinted with permission of the Institute of Mathematical Statistics.