

# 15

## Nonparametric Statistics

---

### CHAPTER OUTLINE

15-1	INTRODUCTION	15-3.3	Paired Observations
15-2	SIGN TEST	15-3.4	Comparison to the $t$ -Test
	15-2.1	Description of the Test	
	15-2.2	Sign Test for Paired Samples	
	15-2.3	Type II Error for the Sign Test	
	15-2.4	Comparison to the $t$ -Test	
15-3	WILCOXON SIGNED-RANK TEST	15-4	WILCOXON RANK-SUM TEST
	15-3.1	Description of the Test	
	15-3.2	Large-Sample Approximation	
		15-4.1	Description of the Test
		15-4.2	Large-Sample Approximation
		15-4.3	Comparison to the $t$ -Test
		15-5	NONPARAMETRIC METHODS IN THE ANALYSIS OF VARIANCE
		15-5.1	Kruskal-Wallis Test
		15-5.2	Rank Transformation

---

### LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Determine situations where nonparametric procedures are better alternatives to the  $t$ -test and ANOVA
  2. Use one- and two-sample nonparametric tests
  3. Use nonparametric alternatives to the single-factor ANOVA
  4. Understand how nonparametric tests compare to the  $t$ -test in terms of relative efficiency
- 

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

## 15-1 INTRODUCTION

Most of the hypothesis-testing and confidence interval procedures discussed in previous chapters are based on the assumption that we are working with random samples from normal populations. Traditionally, we have called these procedures **parametric methods** because they are based on a particular parametric family of distributions—in this case, the normal. Alternately, sometimes we say that these procedures are not *distribution-free* because they depend on the assumption of normality. Fortunately, most of these procedures are relatively insensitive to slight departures from normality. In general, the  $t$ - and  $F$ -tests and the  $t$ -confidence intervals will have actual levels of significance or confidence levels that differ from the nominal or advertised levels chosen by the experimenter, although the difference between the actual and advertised levels is usually fairly small when the underlying population is not too different from the normal.

In this chapter we describe procedures called **nonparametric** and **distribution-free methods**, and we usually make no assumptions about the distribution of the underlying population other than that it is continuous. These procedures have actual level of significance  $\alpha$  or confidence level  $100(1 - \alpha)\%$  for many different types of distributions. These procedures have considerable appeal. One of their advantages is that the data need not be quantitative but can be categorical (such as yes or no, defective or nondefective) or rank data. Another advantage is that nonparametric procedures are usually very quick and easy to perform.

The procedures described in this chapter are competitors of the parametric  $t$ - and  $F$ -procedures described earlier. Consequently, it is important to compare the performance of both parametric and nonparametric methods under the assumptions of both normal and non-normal populations. In general, nonparametric procedures do not utilize all the information provided by the sample. As a result, a nonparametric procedure will be less efficient than the corresponding parametric procedure when the underlying population is normal. This loss of efficiency is reflected by a requirement of a larger sample size for the nonparametric procedure than would be required by the parametric procedure in order to achieve the same power. On the other hand, this loss of efficiency is usually not large, and often the difference in sample size is very small. When the underlying distributions are not close to normal, nonparametric methods have much to offer. They often provide considerable improvement over the normal-theory parametric methods.

Generally, if both parametric and nonparametric methods are applicable to a particular problem, we should use the more efficient parametric procedure. However, the assumptions for the parametric method may be difficult or impossible to justify. For example, the data may be in the form of **ranks**. These situations frequently occur in practice. For instance, a panel of judges may be used to evaluate 10 different formulations of a soft-drink beverage for overall quality, with the “best” formulation assigned rank 1, the “next-best” formulation assigned rank 2, and so forth. It is unlikely that rank data satisfy the normality assumption. Many nonparametric methods involve the analysis of ranks and consequently are ideally suited to this type of problem.

## 15-2 SIGN TEST

### 15-2.1 Description of the Test

The **sign test** is used to test hypotheses about the **median**  $\tilde{\mu}$  of a continuous distribution. The median of a distribution is a value of the random variable  $X$  such that the probability is 0.5 that an observed value of  $X$  is less than or equal to the median, and the probability is 0.5 that an observed value of  $X$  is greater than or equal to the median. That is,  $P(X \leq \tilde{\mu}) = P(X \geq \tilde{\mu}) = 0.5$ .

Since the normal distribution is symmetric, the mean of a normal distribution equals the median. Therefore, the sign test can be used to test hypotheses about the mean of a normal distribution. This is the same problem for which we used the  $t$ -test in Chapter 9. We will discuss the relative merits of the two procedures in Section 15-2.4. Note that, although the  $t$ -test was designed for samples from a normal distribution, the sign test is appropriate for samples from any continuous distribution. Thus, the sign test is a nonparametric procedure.

Suppose that the hypotheses are

$$\begin{aligned} H_0: \tilde{\mu} &= \tilde{\mu}_0 \\ H_1: \tilde{\mu} &< \tilde{\mu}_0 \end{aligned} \quad (15-1)$$

The test procedure is easy to describe. Suppose that  $X_1, X_2, \dots, X_n$  is a random sample from the population of interest. Form the differences

$$X_i - \tilde{\mu}_0, \quad i = 1, 2, \dots, n \quad (15-2)$$

Now if the null hypothesis  $H_0: \tilde{\mu} = \tilde{\mu}_0$  is true, any difference  $X_i - \tilde{\mu}_0$  is equally likely to be positive or negative. An appropriate test statistic is the number of these differences that are positive, say  $R^+$ . Therefore, to test the null hypothesis we are really testing that the number of plus signs is a value of a binomial random variable that has the parameter  $p = 1/2$ . A  $P$ -value for the observed number of plus signs  $r^+$  can be calculated directly from the binomial distribution. For instance, in testing the hypotheses in Equation 15-1, we will reject  $H_0$  in favor of  $H_1$  only if the proportion of plus signs is sufficiently less than  $1/2$  (or equivalently, whenever the observed number of plus signs  $r^+$  is too small). Thus, if the computed  $P$ -value

$$P = P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

is less than or equal to some preselected significance level  $\alpha$ , we will reject  $H_0$  and conclude  $H_1$  is true.

To test the other one-sided hypothesis

$$\begin{aligned} H_0: \tilde{\mu} &= \tilde{\mu}_0 \\ H_1: \tilde{\mu} &> \tilde{\mu}_0 \end{aligned} \quad (15-3)$$

we will reject  $H_0$  in favor of  $H_1$  only if the observed number of plus signs, say  $r^+$ , is large or, equivalently, whenever the observed fraction of plus signs is significantly greater than  $1/2$ . Thus, if the computed  $P$ -value

$$P = P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

is less than  $\alpha$ , we will reject  $H_0$  and conclude that  $H_1$  is true.

The two-sided alternative may also be tested. If the hypotheses are

$$\begin{aligned} H_0: \tilde{\mu} &= \tilde{\mu}_0 \\ H_1: \tilde{\mu} &\neq \tilde{\mu}_0 \end{aligned} \quad (15-4)$$

we should reject  $H_0: \tilde{\mu} = \tilde{\mu}_0$  if the proportion of plus signs is significantly different (either less than or greater than) from  $1/2$ . This is equivalent to the observed number of plus signs  $r^+$  being either sufficiently large or sufficiently small. Thus, if  $r^+ < n/2$  the  $P$ -value is

$$P = 2P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

and if  $r^+ > n/2$  the  $P$ -value is

$$P = 2P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

If the  $P$ -value is less than some preselected level  $\alpha$ , we will reject  $H_0$  and conclude that  $H_1$  is true.

### EXAMPLE 15-1

Montgomery, Peck, and Vining (2001) report on a study in which a rocket motor is formed by binding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two propellant types is an important characteristic. The results of testing 20 randomly selected motors are shown in Table 15-1. We would like to test the hypothesis that the median shear strength is 2000 psi, using  $\alpha = 0.05$ .

This problem can be solved using the eight-step hypothesis-testing procedure introduced in Chapter 9:

1. The parameter of interest is the median of the distribution of propellant shear strength.
2.  $H_0: \tilde{\mu} = 2000$  psi
3.  $H_1: \tilde{\mu} \neq 2000$  psi
4.  $\alpha = 0.05$

Table 15-1 Propellant Shear Strength Data

Observation $i$	Shear Strength $x_i$	Differences $x_i - 2000$	Sign
1	2158.70	+158.70	+
2	1678.15	-321.85	-
3	2316.00	+316.00	+
4	2061.30	+61.30	+
5	2207.50	+207.50	+
6	1708.30	-291.70	-
7	1784.70	-215.30	-
8	2575.10	+575.10	+
9	2357.90	+357.90	+
10	2256.70	+256.70	+
11	2165.20	+165.20	+
12	2399.55	+399.55	+
13	1779.80	-220.20	-
14	2336.75	+336.75	+
15	1765.30	-234.70	-
16	2053.50	+53.50	+
17	2414.40	+414.40	+
18	2200.50	+200.50	+
19	2654.20	+654.20	+
20	1753.70	-246.30	-

5. The test statistic is the observed number of plus differences in Table 15-1, or  $r^+ = 14$ .
6. We will reject  $H_0$  if the  $P$ -value corresponding to  $r^+ = 14$  is less than or equal to  $\alpha = 0.05$ .
7. Computations: Since  $r^+ = 14$  is greater than  $n/2 = 20/2 = 10$ , we calculate the  $P$ -value from

$$\begin{aligned} P &= 2P\left(R^+ \geq 14 \text{ when } p = \frac{1}{2}\right) \\ &= 2 \sum_{r=14}^{20} \binom{20}{r} (0.5)^r (0.5)^{20-r} \\ &= 0.1153 \end{aligned}$$

8. Conclusions: Since  $P = 0.1153$  is not less than  $\alpha = 0.05$ , we cannot reject the null hypothesis that the median shear strength is 2000 psi. Another way to say this is that the observed number of plus signs  $r^+ = 14$  was not large or small enough to indicate that median shear strength is different from 2000 psi at the  $\alpha = 0.05$  level of significance.

It is also possible to construct a table of critical values for the sign test. This table is shown as Appendix Table VII. The use of this table for the two-sided alternative hypothesis in Equation 15-4 is simple. As before, let  $R^+$  denote the number of the differences ( $X_i - \tilde{\mu}_0$ ) that are positive and let  $R^-$  denote the number of these differences that are negative. Let  $R = \min(R^+, R^-)$ . Appendix Table VII presents critical values  $r_\alpha^*$  for the sign test that ensure that  $P$  (type I error) =  $P$  (reject  $H_0$  when  $H_0$  is true) =  $\alpha$  for  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . If the observed value of the test statistic  $r \leq r_\alpha^*$ , the null hypothesis  $H_0: \tilde{\mu} = \tilde{\mu}_0$  should be rejected.

To illustrate how this table is used, refer to the data in Table 15-1 that was used in Example 15-1. Now  $r^+ = 14$  and  $r^- = 6$ ; therefore,  $r = \min(14, 6) = 6$ . From Appendix Table VII with  $n = 20$  and  $\alpha = 0.05$ , we find that  $r_{0.05}^* = 5$ . Since  $r = 6$  is not less than or equal to the critical value  $r_{0.05}^* = 5$ , we cannot reject the null hypothesis that the median shear strength is 2000 psi.

We can also use Appendix Table VII for the sign test when a one-sided alternative hypothesis is appropriate. If the alternative is  $H_1: \tilde{\mu} > \tilde{\mu}_0$ , reject  $H_0: \tilde{\mu} = \tilde{\mu}_0$  if  $r^- \leq r_\alpha^*$ ; if the alternative is  $H_1: \tilde{\mu} < \tilde{\mu}_0$ , reject  $H_0: \tilde{\mu} = \tilde{\mu}_0$  if  $r^+ \leq r_\alpha^*$ . The level of significance of a one-sided test is one-half the value for a two-sided test. Appendix Table VII shows the one-sided significance levels in the column headings immediately below the two-sided levels.

Finally, note that when a test statistic has a discrete distribution such as  $R$  does in the sign test, it may be impossible to choose a critical value  $r_\alpha^*$  that has a level of significance exactly equal to  $\alpha$ . The approach used in Appendix Table VII is to choose  $r_\alpha^*$  to yield an  $\alpha$  that is as close to the advertised significance level  $\alpha$  as possible.

### Ties in the Sign Test

Since the underlying population is assumed to be continuous, there is a zero probability that we will find a “tie”—that is, a value of  $X_i$  exactly equal to  $\tilde{\mu}_0$ . However, this may sometimes happen in practice because of the way the data are collected. When ties occur, they should be set aside and the sign test applied to the remaining data.

### The Normal Approximation

When  $p = 0.5$ , the binomial distribution is well approximated by a normal distribution when  $n$  is at least 10. Thus, since the mean of the binomial is  $np$  and the variance is  $np(1 - p)$ , the distribution of  $R^+$  is approximately normal with mean  $0.5n$  and variance  $0.25n$  whenever  $n$  is moderately large. Therefore, in these cases the null hypothesis  $H_0: \tilde{\mu} = \tilde{\mu}_0$  can be tested using the statistic

$$Z_0 = \frac{R^+ - 0.5n}{0.5\sqrt{n}} \quad (15-5)$$

The two-sided alternative would be rejected if the observed value of the test statistic  $|z_0| > z_{\alpha/2}$ , and the critical regions of the one-sided alternative would be chosen to reflect the sense of the alternative. (If the alternative is  $H_1: \tilde{\mu} > \tilde{\mu}_0$ , reject  $H_0$  if  $z_0 > z_\alpha$ , for example.)

#### EXAMPLE 15-2

We will illustrate the normal approximation procedure by applying it to the problem in Example 15-1. Recall that the data for this example are in Table 15-1. The eight-step procedure follows:

1. The parameter of interest is the median of the distribution of propellant shear strength.
2.  $H_0: \tilde{\mu} = 2000$  psi
3.  $H_1: \tilde{\mu} \neq 2000$  psi
4.  $\alpha = 0.05$
5. The test statistic is

$$z_0 = \frac{r^+ - 0.5n}{0.5\sqrt{n}}$$

6. Since  $\alpha = 0.05$ , we will reject  $H_0$  in favor of  $H_1$  if  $|z_0| > z_{0.025} = 1.96$ .
7. Computations: Since  $r^+ = 14$ , the test statistic is

$$z_0 = \frac{14 - 0.5(20)}{0.5\sqrt{20}} = 1.789$$

8. Conclusions: Since  $z_0 = 1.789$  is not greater than  $z_{0.025} = 1.96$ , we cannot reject the null hypothesis. Thus, our conclusions are identical to those in Example 15-1.

### 15-2.2 Sign Test for Paired Samples

The sign test can also be applied to paired observations drawn from continuous populations. Let  $(X_{1j}, X_{2j}), j = 1, 2, \dots, n$  be a collection of paired observations from two continuous populations, and let

$$D_j = X_{1j} - X_{2j} \quad j = 1, 2, \dots, n$$

be the paired differences. We wish to test the hypothesis that the two populations have a common median, that is, that  $\tilde{\mu}_1 = \tilde{\mu}_2$ . This is equivalent to testing that the median of the differences  $\tilde{\mu}_D = 0$ . This can be done by applying the sign test to the  $n$  observed differences  $d_j$ , as illustrated in the following example.

**EXAMPLE 15-3**

An automotive engineer is investigating two different types of metering devices for an electronic fuel injection system to determine whether they differ in their fuel mileage performance. The system is installed on 12 different cars, and a test is run with each metering device on each car. The observed fuel mileage performance data, corresponding differences, and their signs are shown in Table 15-2. We will use the sign test to determine whether the median fuel mileage performance is the same for both devices using  $\alpha = 0.05$ . The eight-step-procedure follows:

1. The parameters of interest are the median fuel mileage performance for the two metering devices.
2.  $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ , or, equivalently,  $H_0: \tilde{\mu}_D = 0$
3.  $H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2$ , or, equivalently,  $H_1: \tilde{\mu}_D \neq 0$
4.  $\alpha = 0.05$
5. We will use Appendix Table VII to conduct the test, so the test statistic is  $r = \min(r^+, r^-)$ .
6. Since  $\alpha = 0.05$  and  $n = 12$ , Appendix Table VII gives the critical values as  $r_{0.05}^* = 2$ . We will reject  $H_0$  in favor of  $H_1$  if  $r \leq 2$ .
7. Computations: Table 15-2 shows the differences and their signs, and we note that  $r^+ = 8$ ,  $r^- = 4$ , and so  $r = \min(8, 4) = 4$ .
8. Conclusions: Since  $r = 4$  is not less than or equal to the critical value  $r_{0.05}^* = 2$ , we cannot reject the null hypothesis that the two devices provide the same median fuel mileage performance.

**Table 15-2** Performance of Flow Metering Devices

Car	Metering Device		Difference, $d_j$	Sign
	1	2		
1	17.6	16.8	0.8	+
2	19.4	20.0	-0.6	-
3	19.5	18.2	1.3	+
4	17.1	16.4	0.7	+
5	15.3	16.0	-0.7	-
6	15.9	15.4	0.5	+
7	16.3	16.5	-0.2	-
8	18.4	18.0	0.4	+
9	17.3	16.4	0.9	+
10	19.1	20.1	-1.0	-
11	17.8	16.7	1.1	+
12	18.2	17.9	0.3	+

### 15-2.3 Type II Error for the Sign Test

The sign test will control the probability of type I error at an advertised level  $\alpha$  for testing the null hypothesis  $H_0: \tilde{\mu} = \tilde{\mu}$  for any continuous distribution. As with any hypothesis-testing procedure, it is important to investigate the probability of a type II error,  $\beta$ . The test should be able to effectively detect departures from the null hypothesis, and a good measure of this effectiveness is the value of  $\beta$  for departures that are important. A small value of  $\beta$  implies an effective test procedure.

In determining  $\beta$ , it is important to realize not only that a particular value of  $\tilde{\mu}$ , say  $\tilde{\mu}_0 + \Delta$ , must be used but also that the **form** of the underlying distribution will affect the calculations. To illustrate, suppose that the underlying distribution is normal with  $\sigma = 1$  and we are testing the hypothesis  $H_0: \tilde{\mu} = 2$  versus  $H_1: \tilde{\mu} > 2$ . (Since  $\tilde{\mu} = \mu$  in the normal distribution, this is equivalent to testing that the mean equals 2.) Suppose that it is important to detect a departure from  $\tilde{\mu} = 2$  to  $\tilde{\mu} = 3$ . The situation is illustrated graphically in Fig. 15-1(a). When the alternative hypothesis is true ( $H_1: \tilde{\mu} = 3$ ), the probability that the random variable  $X$  is less than or equal to the value 2 is

$$p = P(X \leq 2) = P(Z \leq -1) = \Phi(-1) = 0.1587$$

Suppose we have taken a random sample of size 12. At the  $\alpha = 0.05$  level, Appendix Table VII indicates that we would reject  $H_0: \tilde{\mu} = 2$  if  $r^- \leq r_{0.05}^* = 2$ . Therefore,  $\beta$  is the probability that we do not reject  $H_0: \tilde{\mu} = 2$  when in fact  $\tilde{\mu} = 3$ , or

$$\beta = 1 - \sum_{x=0}^2 \binom{12}{x} (0.1587)^x (0.8413)^{12-x} = 0.2944$$

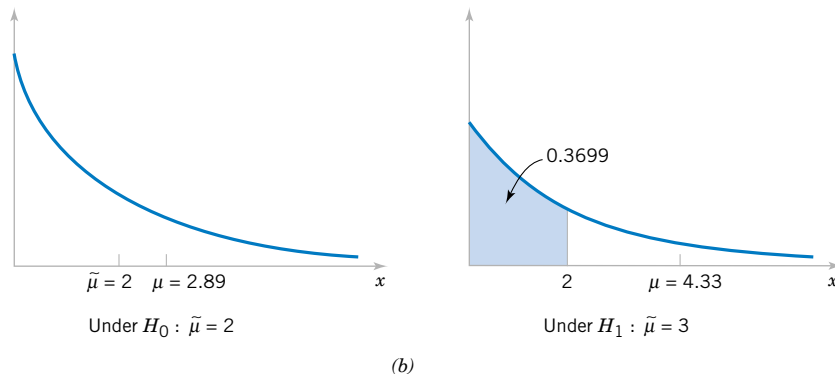
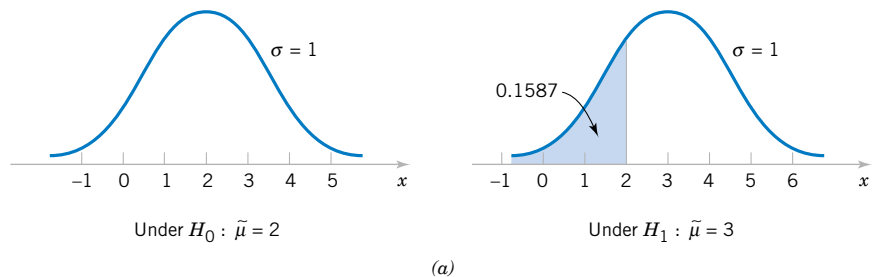


Figure 15-1 Calculation of  $\beta$  for the sign test. (a) Normal distributions. (b) Exponential distributions.



If the distribution of  $X$  had been exponential rather than normal, the situation would be as shown in Fig. 15-1(b), and the probability that the random variable  $X$  is less than or equal to the value  $x = 2$  when  $\tilde{\mu} = 3$  (note that when the median of an exponential distribution is 3, the mean is 4.33) is

$$p = P(X \leq 2) = \int_0^2 \frac{1}{4.33} e^{-\frac{1}{4.33}x} dx = 0.3699$$

In this case,

$$\beta = 1 - \sum_{x=0}^2 \binom{12}{x} (0.3699)^x (0.6301)^{12-x} = 0.8794$$

Thus,  $\beta$  for the sign test depends not only on the alternative value of  $\tilde{\mu}$  but also on the area to the right of the value specified in the null hypothesis under the population probability distribution. This area is highly dependent on the shape of that particular probability distribution.

#### 15-2.4 Comparison to the $t$ -Test

If the underlying population is normal, either the sign test or the  $t$ -test could be used to test  $H_0: \tilde{\mu} = \tilde{\mu}_0$ . The  $t$ -test is known to have the smallest value of  $\beta$  possible among all tests that have significance level  $\alpha$  for the one-sided alternative and for tests with symmetric critical regions for the two-sided alternative, so it is superior to the sign test in the normal distribution case. When the population distribution is symmetric and nonnormal (but with finite mean  $\mu = \tilde{\mu}$ ), the  $t$ -test will have a smaller  $\beta$  (or a higher power) than the sign test, unless the distribution has very heavy tails compared with the normal. Thus, the sign test is usually considered a test procedure for the median rather than as a serious competitor for the  $t$ -test. The **Wilcoxon signed-rank test** discussed in the next section is preferable to the sign test and compares well with the  $t$ -test for symmetric distributions.

#### EXERCISES FOR SECTION 15-2

**15-1.** Ten samples were taken from a plating bath used in an electronics manufacturing process, and the bath pH was determined. The sample pH values are 7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, 7.42. Manufacturing engineering believes that pH has a median value of 7.0. Do the sample data indicate that this statement is correct? Use the sign test with  $\alpha = 0.05$  to investigate this hypothesis. Find the  $P$ -value for this test.

**15-2.** The titanium content in an aircraft-grade alloy is an important determinant of strength. A sample of 20 test coupons reveals the following titanium content (in percent):

8.32, 8.05, 8.93, 8.65, 8.25, 8.46, 8.52, 8.35, 8.36, 8.41, 8.42, 8.30, 8.71, 8.75, 8.60, 8.83, 8.50, 8.38, 8.29, 8.46

The median titanium content should be 8.5%. Use the sign test with  $\alpha = 0.05$  to investigate this hypothesis. Find the  $P$ -value for this test.

**15-3.** The impurity level (in ppm) is routinely measured in an intermediate chemical product. The following data were observed in a recent test:

2.4, 2.5, 1.7, 1.6, 1.9, 2.6, 1.3, 1.9, 2.0, 2.5, 2.6, 2.3, 2.0, 1.8, 1.3, 1.7, 2.0, 1.9, 2.3, 1.9, 2.4, 1.6

Can you claim that the median impurity level is less than 2.5 ppm? State and test the appropriate hypothesis using the sign test with  $\alpha = 0.05$ . What is the  $P$ -value for this test?

**15-4.** Consider the data in Exercise 15-1. Use the normal approximation for the sign test to test  $H_0: \tilde{\mu} = 7.0$  versus  $H_1: \tilde{\mu} \neq 7.0$ . What is the  $P$ -value for this test?

**15-5.** Consider the compressive strength data in Exercise 8-26.

- (a) Use the sign test to investigate the claim that the median strength is at least 2250 psi. Use  $\alpha = 0.05$ .
- (b) Use the normal approximation to test the same hypothesis that you formulated in part (a). What is the  $P$ -value for this test?

**15-6.** Consider the margarine fat content data in Exercise 8-25. Use the sign test to test  $H_0: \tilde{\mu} = 17.0$  versus  $H_1: \tilde{\mu} \neq 17.0$ , with  $\alpha = 0.05$ . Find the  $P$ -value for the test statistic and use this quantity to make your decision.

**15-7.** Consider the data in Exercise 15-2. Use the normal approximation for the sign test to test  $H_0: \tilde{\mu} = 8.5$  versus  $H_1: \tilde{\mu} \neq 8.5$ , with  $\alpha = 0.05$ . What is the  $P$ -value for this test?

**15-8.** Consider the data in Exercise 15-3. Use the normal approximation for the sign test to test  $H_0: \tilde{\mu} = 2.5$  versus  $H_1: \tilde{\mu} < 2.5$ . What is the  $P$ -value for this test?

**15-9.** Two different types of tips can be used in a Rockwell hardness tester. Eight coupons from test ingots of a nickel-based alloy are selected, and each coupon is tested twice, once with each tip. The Rockwell C-scale hardness readings are shown in the following table. Use the sign test with  $\alpha = 0.05$  to determine whether or not the two tips produce equivalent hardness readings.

Coupon	Tip 1	Tip 2
1	63	60
2	52	51
3	58	56
4	60	59
5	55	58
6	57	54
7	53	52
8	59	61

**15-10.** Two different formulations of primer paint can be used on aluminum panels. The drying time of these two formulations is an important consideration in the manufacturing process. Twenty panels are selected; half of each panel is painted with primer 1, and the other half is painted with primer 2. The drying times are observed and reported in the following table. Is there evidence that the median drying times of the two formulations are different? Use the sign test with  $\alpha = 0.01$ .

Panel	Drying Times (in hr)	
	Formulation 1	Formulation 2
1	1.6	1.8
2	1.3	1.5
3	1.5	1.5
4	1.6	1.7
5	1.7	1.6
6	1.9	2.0
7	1.8	2.1
8	1.6	1.7
9	1.4	1.6
10	1.8	1.9
11	1.9	2.0
12	1.8	1.9
13	1.7	1.5
14	1.5	1.7
15	1.6	1.6
16	1.4	1.2
17	1.3	1.6
18	1.6	1.8
19	1.5	1.6
20	1.8	2.0

**15-11.** Use the normal approximation to the sign test for the data in Exercise 15-10. What conclusions can you draw?

**15-12.** The diameter of a ball bearing was measured by 12 inspectors, each using two different kinds of calipers. The results were as follows:

Inspector	Caliper 1	Caliper 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

Is there a significant difference between the medians of the population of measurements represented by the two samples? Use  $\alpha = 0.05$ .

**15-13.** Consider the blood cholesterol data in Exercise 10-39. Use the sign test to determine whether there is any difference between the medians of the two groups of measurements, with  $\alpha = 0.05$ . What practical conclusion would you draw from this study?

**15-14.** Use the normal approximation for the sign test for the data in Exercise 15-12. With  $\alpha = 0.05$ , what conclusions can you draw?

**15-15.** Use the normal approximation to the sign test for the data in Exercise 15-13. With  $\alpha = 0.05$ , what conclusions can you draw?

**15-16.** The distribution time between arrivals in a telecommunication system is exponential, and the system manager wishes to test the hypothesis that  $H_0: \tilde{\mu} = 3.5$  minutes versus  $H_1: \tilde{\mu} > 3.5$  minutes.

- What is the value of the mean of the exponential distribution under  $H_0: \tilde{\mu} = 3.5$ ?
- Suppose that we have taken a sample of  $n = 10$  observations and we observe  $r^- = 3$ . Would the sign test reject  $H_0$  at  $\alpha = 0.05$ ?

(c) What is the type II error probability of this test if  $\tilde{\mu} = 4.5$ ?

**15-17.** Suppose that we take a sample of  $n = 10$  measurements from a normal distribution with  $\sigma = 1$ . We wish to test  $H_0: \mu = 0$  against  $H_1: \mu > 0$ . The normal test statistic is  $Z_0 = \bar{X}/(\sigma/\sqrt{n})$ , and we decide to use a critical region of 1.96 (that is, reject  $H_0$  if  $z_0 \geq 1.96$ ).

- What is  $\alpha$  for this test?
- What is  $\beta$  for this test, if  $\mu = 1$ ?
- If a sign test is used, specify the critical region that gives an  $\alpha$  value consistent with  $\alpha$  for the normal test.
- What is the  $\beta$  value for the sign test, if  $\mu = 1$ ? Compare this with the result obtained in part (b).

**15-18.** Consider the test statistic for the sign test in Exercise 15-9. Find the  $P$ -value for this statistic.

**15-19.** Consider the test statistic for the sign test in Exercise 15-10. Find the  $P$ -value for this statistic. Compare it to the  $P$ -value for the normal approximation test statistic computed in Exercise 15-11.

## 15-3 WILCOXON SIGNED-RANK TEST

The sign test makes use only of the plus and minus signs of the differences between the observations and the median  $\tilde{\mu}_0$  (or the plus and minus signs of the differences between the observations in the paired case). It does not take into account the size or magnitude of these differences. Frank Wilcoxon devised a test procedure that uses both direction (sign) and magnitude. This procedure, now called the **Wilcoxon signed-rank test**, is discussed and illustrated in this section.

The Wilcoxon signed-rank test applies to the case of **symmetric continuous distributions**. Under these assumptions, the mean equals the median, and we can use this procedure to test the null hypothesis that  $m = \mu_0$ . We now show how to do this.

### 15-3.1 Description of the Test

We are interested in testing  $H_0: \mu = \mu_0$  against the usual alternatives. Assume that  $X_1, X_2, \dots, X_n$  is a random sample from a continuous and symmetric distribution with mean (and median)  $\mu$ . Compute the differences  $X_i - \mu_0, i = 1, 2, \dots, n$ . Rank the absolute differences  $|X_i - \mu_0|, i = 1, 2, \dots, n$  in ascending order, and then give the ranks the signs of their corresponding differences. Let  $W^+$  be the sum of the positive ranks and  $W^-$  be the absolute value of the sum of the negative ranks, and let  $W = \min(W^+, W^-)$ . Appendix Table VIII contains critical values of  $W$ , say  $w_\alpha^*$ . If the alternative hypothesis is  $H_1: \mu \neq \mu_0$ , then if the observed value of the statistic  $w \leq w_\alpha^*$ , the null hypothesis  $H_0: \mu = \mu_0$  is rejected. Appendix Table VIII provides significance levels of  $\alpha = 0.10, \alpha = 0.05, \alpha = 0.02, \alpha = 0.01$  for the two-sided test.

For one-sided tests, if the alternative is  $H_1: \mu > \mu_0$ , reject  $H_0: \mu = \mu_0$  if  $w^- \leq w_\alpha^*$ ; and if the alternative is  $H_1: \mu < \mu_0$ , reject  $H_0: \mu = \mu_0$  if  $w^+ \leq w_\alpha^*$ . The significance levels for one-sided tests provided in Appendix Table VIII are  $\alpha = 0.05, 0.025, 0.01, \text{ and } 0.005$ .

**EXAMPLE 15-4**

We will illustrate the Wilcoxon signed-rank test by applying it to the propellant shear strength data from Table 15-1. Assume that the underlying distribution is a continuous symmetric distribution. The eight-step procedure is applied as follows:

1. The parameter of interest is the mean (or median) of the distribution of propellant shear strength.
2.  $H_0: \mu = 2000$  psi
3.  $H_1: \mu \neq 2000$  psi
4.  $\alpha = 0.05$
5. The test statistic is

$$w = \min(w^+, w^-)$$

6. We will reject  $H_0$  if  $w \leq w_{0.05}^* = 52$  from Appendix Table VIII.
7. Computations: The signed ranks from Table 15-1 are shown in the following table:

Observation	Difference $x_i - 2000$	Signed Rank
16	+53.50	+1
4	+61.30	+2
1	+158.70	+3
11	+165.20	+4
18	+200.50	+5
5	+207.50	+6
7	-215.30	-7
13	-220.20	-8
15	-234.70	-9
20	-246.30	-10
10	+256.70	+11
6	-291.70	-12
3	+316.00	+13
2	-321.85	-14
14	+336.75	+15
9	+357.90	+16
12	+399.55	+17
17	+414.40	+18
8	+575.10	+19
19	+654.20	+20

The sum of the positive ranks is  $w^+ = (1 + 2 + 3 + 4 + 5 + 6 + 11 + 13 + 15 + 16 + 17 + 18 + 19 + 20) = 150$ , and the sum of the absolute values of the negative ranks is  $w^- = (7 + 8 + 9 + 10 + 12 + 14) = 60$ . Therefore,

$$w = \min(150, 60) = 60$$

8. Conclusions: Since  $w = 60$  is not less than or equal to the critical value  $w_{0.05} = 52$ , we cannot reject the null hypothesis that the mean (or median, since the population is assumed to be symmetric) shear strength is 2000 psi.

### Ties in the Wilcoxon Signed-Rank Test

Because the underlying population is continuous, ties are theoretically impossible, although they will sometimes occur in practice. If several observations have the same absolute magnitude, they are assigned the average of the ranks that they would receive if they differed slightly from one another.

### 15-3.2 Large-Sample Approximation

If the sample size is moderately large, say  $n > 20$ , it can be shown that  $W^+$  (or  $W^-$ ) has approximately a normal distribution with mean

$$\mu_{W^+} = \frac{n(n+1)}{4}$$

and variance

$$\sigma_{W^+}^2 = \frac{n(n+1)(2n+1)}{24}$$

Therefore, a test of  $H_0: \mu = \mu_0$  can be based on the statistic

$$Z_0 = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \quad (15-6)$$

An appropriate critical region for either the two-sided or one-sided alternative hypotheses can be chosen from a table of the standard normal distribution.

### 15-3.3 Paired Observations

The Wilcoxon signed-rank test can be applied to paired data. Let  $(X_{1j}, X_{2j})$ ,  $j = 1, 2, \dots, n$  be a collection of paired observations from two continuous distributions that differ only with respect to their means. (It is not necessary that the distributions of  $X_1$  and  $X_2$  be symmetric.) This assures that the distribution of the differences  $D_j = X_{1j} - X_{2j}$  is continuous and symmetric. Thus, the null hypothesis is  $H_0: \mu_1 = \mu_2$ , which is equivalent to  $H_0: \mu_D = 0$ . We initially consider the two-sided alternative  $H_1: \mu_1 \neq \mu_2$  (or  $H_1: \mu_D \neq 0$ ).

To use the Wilcoxon signed-rank test, the differences are first ranked in ascending order of their absolute values, and then the ranks are given the signs of the differences. Ties are assigned average ranks. Let  $W^+$  be the sum of the positive ranks and  $W^-$  be the absolute value of the sum of the negative ranks, and  $W = \min(W^+, W^-)$ . If the observed value  $w \leq w_\alpha^*$ , the null hypothesis  $H_0: \mu_1 = \mu_2$  (or  $H_0: \mu_D = 0$ ) is rejected where  $w_\alpha^*$  is chosen from Appendix Table VIII.

For one-sided tests, if the alternative is  $H_1: \mu_1 > \mu_2$  (or  $H_1: \mu_D > 0$ ), reject  $H_0$  if  $w^- \leq w_\alpha^*$ ; and if  $H_1: \mu_1 < \mu_2$  (or  $H_1: \mu_D < 0$ ), reject  $H_0$  if  $w^+ \leq w_\alpha^*$ . Be sure to use the one-sided test significance levels shown in Appendix Table VIII.

#### EXAMPLE 15-5

We will apply the Wilcoxon signed-rank test to the fuel-metering device test data used previously in Example 15-3. The eight-step hypothesis-testing procedure can be applied as follows:

1. The parameters of interest are the mean fuel mileage performance for the two metering devices.

2.  $H_0: \mu_1 = \mu_2$  or, equivalently,  $H_0: \mu_D = 0$
3.  $H_1: \mu_1 \neq \mu_2$  or, equivalently,  $H_1: \mu_D \neq 0$
4.  $\alpha = 0.05$
5. The test statistic is

$$w = \min(w^+, w^-)$$

where  $w^+$  and  $w^-$  are the sums of the positive and negative ranks of the differences in Table 15-2.

6. Since  $\alpha = 0.05$  and  $n = 12$ , Appendix Table VIII gives the critical value as  $w_{0.05}^* = 13$ . We will reject  $H_0: \mu_D = 0$  if  $w \leq 13$ .
7. Computations: Using the data in Table 15-2, we compute the following signed ranks:

Car	Difference	Signed Rank
7	-0.2	-1
12	0.3	2
8	0.4	3
6	0.5	4
2	-0.6	-5
4	0.7	6.5
5	-0.7	-6.5
1	0.8	8
9	0.9	9
10	-1.0	-10
11	1.1	11
3	1.3	12

Note that  $w^+ = 55.5$  and  $w^- = 22.5$ . Therefore,

$$w = \min(55.5, 22.5) = 22.5$$

8. Conclusions: Since  $w = 22.5$  is not less than or equal to  $w_{0.05}^* = 13$ , we cannot reject the null hypothesis that the two metering devices produce the same mileage performance.

### 15-3.4 Comparison to the $t$ -Test

When the underlying population is normal, either the  $t$ -test or the Wilcoxon signed-rank test can be used to test hypotheses about  $\mu$ . As mentioned earlier, the  $t$ -test is the best test in such situations in the sense that it produces a minimum value of  $\beta$  for all tests with significance level  $\alpha$ . However, since it is not always clear that the normal distribution is appropriate, and since in many situations it is inappropriate, it is of interest to compare the two procedures for both normal and nonnormal populations.

Unfortunately, such a comparison is not easy. The problem is that  $\beta$  for the Wilcoxon signed-rank test is very difficult to obtain, and  $\beta$  for the  $t$ -test is difficult to obtain for nonnormal distributions. Because type II error comparisons are difficult, other measures of comparison have been developed. One widely used measure is **asymptotic relative efficiency** (ARE).

The ARE of one test relative to another is the limiting ratio of the sample sizes necessary to obtain identical error probabilities for the two procedures. For example, if the ARE of one test relative to the competitor is 0.5, when sample sizes are large, the first test will require twice as large a sample as the second one to obtain similar error performance. While this does not tell us anything for small sample sizes, we can say the following:

1. For normal populations, the ARE of the Wilcoxon signed-rank test relative to the  $t$ -test is approximately 0.95.
2. For nonnormal populations, the ARE is at least 0.86, and in many cases it will exceed unity.

Although these are large-sample results, we generally conclude that the Wilcoxon signed-rank test will never be much worse than the  $t$ -test and that in many cases where the population is non-normal it may be superior. Thus, the Wilcoxon signed-rank test is a useful alternate to the  $t$ -test.

### EXERCISES FOR SECTION 15-3

**15-20.** Consider the data in Exercise 15-1 and assume that the distribution of pH is symmetric and continuous. Use the Wilcoxon signed-rank test with  $\alpha = 0.05$  to test the hypothesis  $H_0: \mu = 7$  against  $H_1: \mu \neq 7$ .

**15-21.** Consider the data in Exercise 15-2. Suppose that the distribution of titanium content is symmetric and continuous. Use the Wilcoxon signed-rank test with  $\alpha = 0.05$  to test the hypothesis  $H_0: \mu = 8.5$  versus  $H_1: \mu \neq 8.5$ .

**15-22.** Consider the data in Exercise 15-2. Use the large-sample approximation for the Wilcoxon signed-rank test to test the hypothesis  $H_0: \mu = 8.5$  versus  $H_1: \mu \neq 8.5$ . Use  $\alpha = 0.05$ . Assume that the distribution of titanium content is continuous and symmetric.

**15-23.** Consider the data in Exercise 15-3. Use the Wilcoxon signed-rank test to test the hypothesis  $H_0: \mu = 2.5$  ppm versus  $H_1: \mu < 2.5$  ppm with  $\alpha = 0.05$ . Assume that the distribution of impurity level is continuous and symmetric.

**15-24.** Consider the Rockwell hardness test data in Exercise 15-9. Assume that both distributions are continuous and use the Wilcoxon signed-rank test to test that the mean difference in hardness readings between the two tips is zero. Use  $\alpha = 0.05$ .

**15-25.** Consider the paint drying time data in Exercise 15-10. Assume that both populations are continuous, and use the Wilcoxon signed-rank test to test that the difference in mean drying times between the two formulations is zero. Use  $\alpha = 0.01$ .

**15-26.** Apply the Wilcoxon signed-rank test to the measurement data in Exercise 15-12. Use  $\alpha = 0.05$  and assume that the two distributions of measurements are continuous.

**15-27.** Apply the Wilcoxon signed-rank test to the blood cholesterol data from Exercise 10-39. Use  $\alpha = 0.05$  and assume that the two distributions are continuous.

## 15-4 WILCOXON RANK-SUM TEST

Suppose that we have two independent continuous populations  $X_1$  and  $X_2$  with means  $\mu_1$  and  $\mu_2$ . Assume that the distributions of  $X_1$  and  $X_2$  have the same shape and spread and differ only (possibly) in their locations. The Wilcoxon rank-sum test can be used to test the hypothesis  $H_0: \mu_1 = \mu_2$ . This procedure is sometimes called the Mann-Whitney test, although the Mann-Whitney test statistic is usually expressed in a different form.

### 15-4.1 Description of the Test

Let  $X_{11}, X_{12}, \dots, X_{1n_1}$  and  $X_{21}, X_{22}, \dots, X_{2n_2}$  be two independent random samples of sizes  $n_1 \leq n_2$  from the continuous populations  $X_1$  and  $X_2$  described earlier. We wish to test the hypotheses

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

The test procedure is as follows. Arrange all  $n_1 + n_2$  observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical), use the mean of the ranks that would have been assigned if the observations differed.

Let  $W_1$  be the sum of the ranks in the smaller sample (1), and define  $W_2$  to be the sum of the ranks in the other sample. Then,

$$W_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - W_1 \quad (15-7)$$

Now if the sample means do not differ, we will expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we will conclude that the means are not equal.

Appendix Table IX contains the critical value of the rank sums for  $\alpha = 0.05$  and  $\alpha = 0.01$  assuming the two-sided alternative above. Refer to Appendix Table IX with the appropriate sample sizes  $n_1$  and  $n_2$ , and the critical value  $w_\alpha$  can be obtained. The null  $H_0: \mu_1 = \mu_2$  is rejected in favor of  $H_1: \mu_1 \neq \mu_2$  if either of the observed values  $w_1$  or  $w_2$  is less than or equal to the tabulated critical value  $w_\alpha$ .

The procedure can also be used for one-sided alternatives. If the alternative is  $H_1: \mu_1 < \mu_2$ , reject  $H_0$  if  $w_1 \leq w_\alpha$ ; for  $H_1: \mu_1 > \mu_2$ , reject  $H_0$  if  $w_2 \leq w_\alpha$ . For these one-sided tests, the tabulated critical values  $w_\alpha$  correspond to levels of significance of  $\alpha = 0.025$  and  $\alpha = 0.005$ .

#### EXAMPLE 15-6

The mean axial stress in tensile members used in an aircraft structure is being studied. Two alloys are being investigated. Alloy 1 is a traditional material, and alloy 2 is a new aluminum-lithium alloy that is much lighter than the standard material. Ten specimens of each alloy type are tested, and the axial stress is measured. The sample data are assembled in Table 15-3. Using  $\alpha = 0.05$ , we wish to test the hypothesis that the means of the two stress distributions are identical.

We will apply the eight-step hypothesis-testing procedure to this problem:

1. The parameters of interest are the means of the two distributions of axial stress.
2.  $H_0: \mu_1 = \mu_2$
3.  $H_1: \mu_1 \neq \mu_2$
4.  $\alpha = 0.05$
5. We will use the Wilcoxon rank-sum test statistic in Equation 15-7,

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

6. Since  $\alpha = 0.05$  and  $n_1 = n_2 = 10$ , Appendix Table IX gives the critical value as  $w_{0.05} = 78$ . If either  $w_1$  or  $w_2$  is less than or equal to  $w_{0.05} = 78$ , we will reject  $H_0: \mu_1 = \mu_2$ .

**Table 15-3** Axial Stress for Two Aluminum-Lithium Alloys

Alloy 1		Alloy 2	
238 psi	3254 psi	3261 psi	3248 psi
3195	3229	3187	3215
3246	3225	3209	3226
3190	3217	3212	3240
3204	3241	3258	3234



7. Computations: The data from Table 15-3 are analyzed in ascending order and ranked as follows:

Alloy Number	Axial Stress	Rank
2	3187 psi	1
1	3190	2
1	3195	3
1	3204	4
2	3209	5
2	3212	6
2	3215	7
1	3217	8
1	3225	9
2	3226	10
1	3229	11
2	3234	12
1	3238	13
2	3240	14
1	3241	15
1	3246	16
2	3248	17
1	3254	18
2	3258	19
2	3261	20

The sum of the ranks for alloy 1 is

$$w_1 = 2 + 3 + 4 + 8 + 9 + 11 + 13 + 15 + 16 + 18 = 99$$

and for alloy 2

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 = \frac{(10 + 10)(10 + 10 + 1)}{2} - 99 = 111$$

8. Conclusions: Since neither  $w_1$  nor  $w_2$  is less than or equal to  $w_{0.05} = 78$ , we cannot reject the null hypothesis that both alloys exhibit the same mean axial stress.

### 15-4.2 Large-Sample Approximation

When both  $n_1$  and  $n_2$  are moderately large, say, greater than 8, the distribution of  $w_1$  can be well approximated by the normal distribution with mean

$$\mu_{w_1} = \frac{n_1(n_1 + n_2 + 1)}{2}$$

and variance

$$\sigma_{w_1}^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Therefore, for  $n_1$  and  $n_2 > 8$ , we could use

$$Z_0 = \frac{W_1 - \mu_{W_1}}{\sigma_{W_1}} \tag{15-8}$$

as a statistic, and the appropriate critical region is  $|z_0| > z_{\alpha/2}$ ,  $z_0 > z_\alpha$ , or  $z_0 < -z_\alpha$ , depending on whether the test is a two-tailed, upper-tail, or lower-tail test.

### 15-4.3 Comparison to the $t$ -Test

In Section 15-3.4 we discussed the comparison of the  $t$ -test with the Wilcoxon signed-rank test. The results for the two-sample problem are identical to the one-sample case. That is, when the normality assumption is correct, the Wilcoxon rank-sum test is approximately 95% as efficient as the  $t$ -test in large samples. On the other hand, regardless of the form of the distributions, the Wilcoxon rank-sum test will always be at least 86% as efficient. The efficiency of the Wilcoxon test relative to the  $t$ -test is usually high if the underlying distribution has heavier tails than the normal, because the behavior of the  $t$ -test is very dependent on the sample mean, which is quite unstable in heavy-tailed distributions.

### EXERCISES FOR SECTION 15-4

**15-28.** An electrical engineer must design a circuit to deliver the maximum amount of current to a display tube to achieve sufficient image brightness. Within her allowable design constraints, she has developed two candidate circuits and tests prototypes of each. The resulting data (in microamperes) are as follows:

**Circuit 1:** 251, 255, 258, 257, 250, 251, 254, 250, 248

**Circuit 2:** 250, 253, 249, 256, 259, 252, 260, 251

Use the Wilcoxon rank-sum test to test  $H_0: \mu_1 = \mu_2$  against the alternative  $H_1: \mu_1 > \mu_2$ . Use  $\alpha = 0.025$ .

**15-29.** One of the authors travels regularly to Seattle, Washington. He uses either Delta or Alaska. Flight delays are sometimes unavoidable, but he would be willing to give most of his business to the airline with the best on-time arrival record. The number of minutes that his flight arrived late for the last six trips on each airline follows. Is there evidence that either airline has superior on-time arrival performance? Use  $\alpha = 0.01$  and the Wilcoxon rank-sum test.

**Delta:** 13, 10, 1, -4, 0, 9 (minutes late)

**Alaska:** 15, 8, 3, -1, -2, 4 (minutes late)

**15-30.** The manufacturer of a hot tub is interested in testing two different heating elements for his product. The element that produces the maximum heat gain after 15 minutes would be preferable. He obtains 10 samples of each heating unit and tests each one. The heat gain after 15 minutes (in °F) follows.

Is there any reason to suspect that one unit is superior to the other? Use  $\alpha = 0.05$  and the Wilcoxon rank-sum test.

**Unit 1:** 25, 27, 29, 31, 30, 26, 24, 32, 33, 38

**Unit 2:** 31, 33, 32, 35, 34, 29, 38, 35, 37, 30

**15-31.** Use the normal approximation for the Wilcoxon rank-sum test for the problem in Exercise 15-28. Assume that  $\alpha = 0.05$ . Find the approximate  $P$ -value for this test statistic.

**15-32.** Use the normal approximation for the Wilcoxon rank-sum test for the heat gain experiment in Exercise 15-30. Assume that  $\alpha = 0.05$ . What is the approximate  $P$ -value for this test statistic?

**15-33.** Consider the chemical etch rate data in Exercise 10-21. Use the Wilcoxon rank-sum test to investigate the claim that the mean etch rate is the same for both solutions. If  $\alpha = 0.05$ , what are your conclusions?

**15-34.** Use the Wilcoxon rank-sum test for the pipe deflection temperature experiment described in Exercise 10-20. If  $\alpha = 0.05$ , what are your conclusions?

**15-35.** Use the normal approximation for the Wilcoxon rank-sum test for the problem in Exercise 10-21. Assume that  $\alpha = 0.05$ . Find the approximate  $P$ -value for this test.

**15-36.** Use the normal approximation for the Wilcoxon rank-sum test for the problem in Exercise 10-20. Assume that  $\alpha = 0.05$ . Find the approximate  $P$ -value for this test.

## 15-5 NONPARAMETRIC METHODS IN THE ANALYSIS OF VARIANCE

### 15-5.1 Kruskal-Wallis Test

The single-factor analysis of variance model developed in Chapter 13 for comparing  $a$  population means is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n_i \end{cases} \quad (15-9)$$

In this model, the error terms  $\epsilon_{ij}$  are assumed to be normally and independently distributed with mean zero and variance  $\sigma^2$ . The assumption of normality led directly to the  $F$ -test described in Chapter 13. The Kruskal-Wallis test is a nonparametric alternative to the  $F$ -test; it requires only that the  $\epsilon_{ij}$  have the same continuous distribution for all factor levels  $i = 1, 2, \dots, a$ .

Suppose that  $N = \sum_{i=1}^a n_i$  is the total number of observations. Rank all  $N$  observations from smallest to largest, and assign the smallest observation rank 1, the next smallest rank 2, . . . , and the largest observation rank  $N$ . If the null hypothesis

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

is true, the  $N$  observations come from the same distribution, and all possible assignments of the  $N$  ranks to the  $a$  samples are equally likely, we would expect the ranks 1, 2, . . . ,  $N$  to be mixed throughout the  $a$  samples. If, however, the null hypothesis  $H_0$  is false, some samples will consist of observations having predominantly small ranks, while other samples will consist of observations having predominantly large ranks. Let  $R_{ij}$  be the rank of observation  $Y_{ij}$ , and let  $R_{i\cdot}$  and  $\bar{R}_i$  denote the total and average of the  $n_i$  ranks in the  $i$ th treatment. When the null hypothesis is true,

$$E(R_{ij}) = \frac{N + 1}{2}$$

and

$$E(\bar{R}_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} E(R_{ij}) = \frac{N + 1}{2}$$

The Kruskal-Wallis test statistic measures the degree to which the actual observed average ranks  $\bar{R}_i$  differ from their expected value  $(N + 1)/2$ . If this difference is large, the null hypothesis  $H_0$  is rejected. The test statistic is

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^a n_i \left( \bar{R}_i - \frac{N + 1}{2} \right)^2 \quad (15-10)$$

An alternative computing formula that is occasionally more convenient is

$$H = \frac{12}{N(N+1)} \sum_{i=1}^a \frac{R_i^2}{n_i} - 3(N+1) \quad (15-11)$$

We would usually prefer Equation 15-11 to Equation 15-10 because it involves the rank totals rather than the averages.

The null hypothesis  $H_0$  should be rejected if the sample data generate a large value for  $H$ . The null distribution for  $H$  has been obtained by using the fact that under  $H_0$  each possible assignment of ranks to the  $a$  treatments is equally likely. Thus, we could enumerate all possible assignments and count the number of times each value of  $H$  occurs. This has led to tables of the critical values of  $H$ , although most tables are restricted to small sample sizes  $n_i$ . In practice, we usually employ the following large-sample approximation. Whenever  $H_0$  is true and either

$$\begin{aligned} a = 3 \quad \text{and} \quad n_i \geq 6 \quad \text{for } i = 1, 2, 3 \\ a > 3 \quad \text{and} \quad n_i \geq 5 \quad \text{for } i = 1, 2, \dots, a \end{aligned}$$

$H$  has approximately a chi-square distribution with  $a - 1$  degrees of freedom. Since large values of  $H$  imply that  $H_0$  is false, we will reject  $H_0$  if the observed value

$$h \geq \chi_{\alpha, a-1}^2$$

The test has approximate significance level  $\alpha$ .

#### Ties in the Kruskal-Wallis Test

When observations are tied, assign an average rank to each of the tied observations. When there are ties, we should replace the test statistic in Equation 15-11 by

$$H = \frac{1}{S^2} \left[ \sum_{i=1}^a \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \quad (15-12)$$

where  $n_i$  is the number of observations in the  $i$ th treatment,  $N$  is the total number of observations, and

$$S^2 = \frac{1}{N-1} \left[ \sum_{i=1}^a \sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right] \quad (15-13)$$

Note that  $S^2$  is just the variance of the ranks. When the number of ties is moderate, there will be little difference between Equations 15-11 and 15-12 and the simpler form (Equation 15-11) may be used.

#### EXAMPLE 15-7

Montgomery (2001) presented data from an experiment in which five different levels of cotton content in a synthetic fiber were tested to determine whether cotton content has any effect on fiber tensile strength. The sample data and ranks from this experiment are shown in Table 15-4. We will apply the Kruskal-Wallis test to these data, using  $\alpha = 0.01$ .

Table 15-4 Data and Ranks for the Tensile Testing Experiment

Percentage of Cotton							$r_i$
15	$y_{1j}$	7	7	15	11	9	
ranks	$r_{1j}$	2.0	2.0	12.5	7.0	4.0	27.5
20	$y_{2j}$	12	17	12	18	18	
ranks	$r_{2j}$	9.5	14.0	9.5	16.5	16.5	66.0
25	$y_{3j}$	14	18	18	19	19	
ranks	$r_{3j}$	11.0	16.5	16.5	20.5	20.5	85.0
30	$y_{4j}$	19	25	22	19	23	
ranks	$r_{4j}$	20.5	25.0	23.0	20.5	24.0	113.0
35	$y_{5j}$	7	10	11	15	11	
ranks	$r_{5j}$	2.0	5.0	7.0	12.5	7.0	33.5

Since there is a fairly large number of ties, we use Equation 15-12 as the test statistic. From Equation 15-13 we find

$$\begin{aligned}
 s^2 &= \frac{1}{N-1} \left[ \sum_{i=1}^a \sum_{j=1}^{n_i} r_{ij}^2 - \frac{N(N+1)^2}{4} \right] \\
 &= \frac{1}{24} \left[ 5510 - \frac{25(26)^2}{2} \right] \\
 &= 53.54
 \end{aligned}$$

and the test statistic is

$$\begin{aligned}
 h &= \frac{1}{s^2} \left[ \sum_{i=1}^a \frac{r_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] \\
 &= \frac{1}{53.54} \left[ 5245.7 - \frac{25(26)^2}{4} \right] \\
 &= 19.06
 \end{aligned}$$

Since  $h > \chi_{0.01,4}^2 = 13.28$ , we would reject the null hypothesis and conclude that treatments differ. This same conclusion is given by the usual analysis of variance  $F$ -test.

## 15-5.2 Rank Transformation

The procedure used in the previous section whereby the observations are replaced by their ranks is called the **rank transformation**. It is a very powerful and widely useful technique. If we were to apply the ordinary  $F$ -test to the ranks rather than to the original data, we would obtain

$$F_0 = \frac{H/(a-1)}{(N-1-H)/(N-a)}$$

as the test statistic. Note that as the Kruskal-Wallis statistic  $H$  increases or decreases,  $F_0$  also increases or decreases. Now, since the distribution of  $F_0$  is approximated by the  $F$ -distribution,

the Kruskal-Wallis test is approximately equivalent to applying the usual analysis of variance to the ranks.

The rank transformation has wide applicability in experimental design problems for which no nonparametric alternative to the analysis of variance exists. If the data are ranked and the ordinary  $F$ -test is applied, an approximate procedure results, but one that has good statistical properties. When we are concerned about the normality assumption or the effect of outliers or “wild” values, we recommend that the usual analysis of variance be performed on both the original data and the ranks. When both procedures give similar results, the analysis of variance assumptions are probably satisfied reasonably well, and the standard analysis is satisfactory. When the two procedures differ, the rank transformation should be preferred since it is less likely to be distorted by nonnormality and unusual observations. In such cases, the experimenter may want to investigate the use of transformations for nonnormality and examine the data and the experimental procedure to determine whether outliers are present and why they have occurred.

**EXERCISES FOR SECTION 15-5**

**15-37.** Montgomery (2001) presented the results of an experiment to compare four different mixing techniques on the tensile strength of portland cement. The results are shown in the following table. Is there any indication that mixing technique affects the strength? Use  $\alpha = 0.05$ .

Mixing Technique	Tensile Strength (lb/in. <sup>2</sup> )				
1	3129	3000	2865	2890	
2	3200	3000	2975	3150	
3	2800	2900	2985	3050	
4	2600	2700	2600	2765	

**15-38.** An article in the *Quality Control Handbook*, 3rd edition (McGraw-Hill, 1962) presents the results of an experiment performed to investigate the effect of three different conditioning methods on the breaking strength of cement briquettes. The data are shown in the following table. Using  $\alpha = 0.05$ , is there any indication that conditioning method affects breaking strength?

Conditioning Method	Breaking Strength (lb/in. <sup>2</sup> )				
1	553	550	568	541	537
2	553	599	579	545	540
3	492	530	528	510	571

**15-39.** In *Statistics for Research* (John Wiley & Sons, 1983), S. Dowdy and S. Wearden presented the results of an experiment to measure the performance of hand-held chain saws. The experimenters measured the kickback angle through which the saw is deflected when it begins to cut a 3-inch stock synthetic board. Shown in the following table are

deflection angles for five saws chosen at random from each of four different manufacturers. Is there any evidence that the manufacturers’ products differ with respect to kickback angle? Use  $\alpha = 0.01$ .

Manufacturer	Kickback Angle				
A	42	17	24	39	43
B	28	50	44	32	61
C	57	45	48	41	54
D	29	40	22	34	30

**15-40.** Consider the data in Exercise 13-2. Use the Kruskal-Wallis procedure with  $\alpha = 0.05$  to test for differences between mean uniformity at the three different gas flow rates.

**15-41.** Find the approximate  $P$ -value for the test statistic computed in Exercise 15-37.

**15-42.** Find the approximate  $P$ -value for the test statistic computed in Exercise 15-40.

**Supplemental Exercises**

**15-43.** The surface finish of 10 metal parts produced in a grinding process is as follows: (in microinches): 10.32, 9.68, 9.92, 10.10, 10.20, 9.87, 10.14, 9.74, 9.80, 10.26. Do the data support the claim that the median value of surface finish is 10 microinches? Use the sign test with  $\alpha = 0.05$ . What is the  $P$ -value for this test?

**15-44.** Use the normal approximation for the sign test for the problem in Exercise 15-43. Find the  $P$ -value for this test. What are your conclusions if  $\alpha = 0.05$ ?

**15-45.** Fluoride emissions (in ppm) from a chemical plant are monitored routinely. The following are 15 observations

based on air samples taken randomly during one month of production: 7, 3, 4, 2, 5, 6, 9, 8, 7, 3, 4, 4, 3, 2, 6. Can you claim that the median fluoride impurity level is less than 6 ppm? State and test the appropriate hypotheses using the sign test with  $\alpha = 0.05$ . What is the  $P$ -value for this test?

**15-46.** Use the normal approximation for the sign test for the problem in Exercise 15-45. What is the  $P$ -value for this test?

**15-47.** Consider the data in Exercise 10-42. Use the sign test with  $\alpha = 0.05$  to determine whether there is a difference in median impurity readings between the two analytical tests.

**15-48.** Consider the data in Exercise 15-43. Use the Wilcoxon signed-rank test for this problem with  $\alpha = 0.05$ . What hypotheses are being tested in this problem?

**15-49.** Consider the data in Exercise 15-45. Use the Wilcoxon signed-rank test for this problem with  $\alpha = 0.05$ . What conclusions can you draw? Does the hypothesis you are testing now differ from the one tested originally in Exercise 15-45?

**15-50.** Use the Wilcoxon signed-rank test with  $\alpha = 0.05$  for the diet-modification experiment described in Exercise 10-41. State carefully the conclusions that you can draw from this experiment.

**15-51.** Use the Wilcoxon rank-sum test with  $\alpha = 0.01$  for the fuel-economy study described in Exercise 10-83. What

conclusions can you draw about the difference in mean mileage performance for the two vehicles in this study?

**15-52.** Use the large-sample approximation for the Wilcoxon rank-sum test for the fuel-economy data in Exercise 10-83. What conclusions can you draw about the difference in means if  $\alpha = 0.01$ ? Find the  $P$ -value for this test.

**15-53.** Use the Wilcoxon rank-sum test with  $\alpha = 0.025$  for the fill-capability experiment described in Exercise 10-85. What conclusions can you draw about the capability of the two fillers?

**15-54.** Use the large-sample approximation for the Wilcoxon rank-sum test with  $\alpha = 0.025$  for the fill-capability experiment described in Exercise 10-85. Find the  $P$ -value for this test. What conclusions can you draw?

**15-55.** Consider the contact resistance experiment in Exercise 13-31. Use the Kruskal-Wallis test to test for differences in mean contact resistance among the three alloys. If  $\alpha = 0.01$ , what are your conclusions? Find the  $P$ -value for this test.

**15-56.** Consider the experiment described in Exercise 13-28. Use the Kruskal-Wallis test for this experiment with  $\alpha = 0.05$ . What conclusions would you draw? Find the  $P$ -value for this test.

**15-57.** Consider the bread quality experiment in Exercise 13-35. Use the Kruskal-Wallis test with  $\alpha = 0.01$  to analyze the data from this experiment. Find the  $P$ -value for this test. What conclusions can you draw?

### MIND-EXPANDING EXERCISES

**15-58.** For the large-sample approximation to the Wilcoxon signed-rank test, derive the mean and standard deviation of the test statistic used in the procedure.

**15-59. Testing for Trends.** A turbocharger wheel is manufactured using an investment-casting process. The shaft fits into the wheel opening, and this wheel opening is a critical dimension. As wheel wax patterns are formed, the hard tool producing the wax patterns wears. This may cause growth in the wheel-opening dimension. Ten wheel-opening measurements, in time order of production, are 4.00 (millimeters), 4.02, 4.03, 4.01, 4.00, 4.03, 4.04, 4.02, 4.03, 4.03.

- (a) Suppose that  $p$  is the probability that observation  $X_{i+5}$  exceeds observation  $X_i$ . If there is no upward or downward trend,  $X_{i+5}$  is no more or less likely to exceed  $X_i$  or lie below  $X_i$ . What is the value of  $p$ ?
- (b) Let  $V$  be the number of values of  $i$  for which  $X_{i+5} > X_i$ . If there is no upward or downward trend

in the measurements, what is the probability distribution of  $V$ ?

- (c) Use the data above and the results of parts (a) and (b) to test  $H_0$ : there is no trend, versus  $H_1$ : there is upward trend. Use  $\alpha = 0.05$ .

Note that this test is a modification of the sign test. It was developed by Cox and Stuart.

**15-60.** Consider the Wilcoxon signed-rank test, and suppose that  $n = 5$ . Assume that  $H_0: \mu = \mu_0$  is true.

- (a) How many different sequences of signed ranks are possible? Enumerate these sequences.
- (b) How many different values of  $W^+$  are there? Find the probability associated with each value of  $W^+$ .
- (c) Suppose that we define the critical region of the test to be to reject  $H_0$  if  $w^+ > w_\alpha^*$  and  $w_\alpha^* = 13$ . What is the approximate  $\alpha$  level of this test?
- (d) Does this exercise show how the critical values for the Wilcoxon signed-rank test were developed? Explain.

**IMPORTANT TERMS AND CONCEPTS**

---

*In the E-book, click on any term or concept below to go to that subject.*

Asymptotic relative efficiency

Kruskal-Wallis test

Nonparametric or distribution free procedures

Normal approximation to nonparametric tests

Paired observations

Ranks

Rank transformation in ANOVA

Sign test

Tests on the mean of a symmetric continuous distribution

Tests on the median of a continuous distribution

Wilcoxon rank-sum test

Wilcoxon signed-rank test