CHAPTER 17

Nonparametric Statistics

LEARNING OBJECTIVES

This chapter presents several nonparametric statistics that can be used to analyze data specifically, thereby enabling you to:

1. Use both the small-sample and large-sample runs tests to determine whether the order of observations in a sample is random.
2. Use both the small-sample and large-sample cases of the Mann-Whitney *U* test to determine if there is a difference in two independent populations.
3. Use both the small-sample and large-sample cases of the Wilcoxon matched-pairs signed rank test to compare the difference in two related samples.
4. Use the Kruskal-Wallis test to determine whether samples come from the same or different populations.
5. Use the Friedman test to determine whether different treatment levels come from the same population when a blocking variable is available.
6. Use Spearman's rank correlation to analyze the degree of association of two variables.

How Is the Doughnut Business?

By investing $5,000, William Rosenberg founded the Industrial Luncheon Services company in 1946 to deliver meals and coffee break snacks to customers in suburban Boston, Massachusetts. Building on his success in this venture, Rosenberg opened his first coffee and doughnut shop called the "Open Kettle" in 1948. In 1950, Rosenberg changed the name of his shop, located in Quincy, Massachusetts, to Dunkin' Donuts, and thus, the first Dunkin' Donuts shop was established. The first Dunkin' Donuts franchise was awarded in 1955, and by 1963, there were 100 Dunkin' Donuts shops. In 1970, the first overseas Dunkin' Donuts shop was opened in Japan, and by 1979, there were 1000 Dunkin' Donuts shops.

Today, there are over 8800 Dunkin' Donuts worldwide in 31 countries. In the United States alone, there are over 6400 Dunkin' Donuts locations. Dunkin' Donuts is the world's largest coffee and baked goods chain, serving more than 3 million customers per day. Dunkin' Donuts sells 52 varieties of doughnuts and more than a dozen coffee beverages as well as an array of bagels, breakfast sandwiches, and other baked goods. Presently, Dunkin' Donuts is a brand of Dunkin' Brands, Inc., based in Canton, Massachusetts.

Suppose researchers at Dunkin' Donuts are studying several manufacturing and marketing questions in an effort to improve the consistency of their products and understand their market. Manufacturing engineers are concerned that the various machines produce a consistent doughnut size. In an effort to test this issue, four machines are selected for a study. Each machine is set to produce a doughnut that is supposed to be about 7.62 cm (3 inches) in diameter. A random sample of doughnuts is taken from each machine and the diameters of the doughnuts are measured. The result is the data shown as follows:

Machine 1	Machine 2	Machine 3	Machine 4
7.58	7.41	7.56	7.72
7.52	7.44	7.55	7.65
7.50	7.42	7.50	7.67
7.52	7.38	7.58	7.70
7.48	7.45	7.53	7.69
	7.40		7.71
			7.73

Suppose Dunkin' Donuts implements a national advertising campaign in the United States. Marketing researchers want to determine whether the campaign has increased the number of doughnuts sold at various outlets around the country. Ten stores are randomly selected and the number of doughnuts sold between 8 and 9 A.M. on a Tuesday is measured both before and after the campaign is implemented. The data follow:

Outlet	Before	After
1	301	374
2	198	187
3	278	332
4	205	212
5	249	243
6	410	478
7	360	386
8	124	141
9	253	251
10	190	264

Do bigger stores have greater sales? To test this question, suppose sales data were gathered from seven Dunkin' Donuts stores along with store size. These figures are used to rank the seven stores on each variable. The ranked data follow.

Store	Sales Rank	Size Rank
1	6	7
2	2	2
3	3	6
4	7	5
5	5	4
6	1	1
7	4	3

Managerial and Statistical Questions

1. The manufacturing researchers who are testing to determine whether there is a difference in the size of doughnuts by machine want to run a one-way ANOVA, but they have serious doubts that the ANOVA assumptions can be met by these data. Is it still possible to analyze the data using statistics?
2. The market researchers are uncertain that normal distribution assumptions underlying the matched-pairs *t* test can be met with the number of doughnuts data. How can the before-and-after data still be used to test the effectiveness of the advertisements?

3. If the sales and store size data are given as ranks, how do we compute a correlation to answer the research question about the relationship of sales and store size? The Pearson product-moment correlation coefficient requires at least interval-level data, and these data are given as ordinal level.

Source: Adapted from information presented on the Dunkin' Donuts' Web site at: http://www.dunkindonuts.com/aboutus./company/. Please note that the data set forth in the problem is fictional, was not supplied by Dunkin' Donuts, and does not necessarily represent Dunkin' Donuts' experience.

Except for the chi-square analyses presented in Chapter 16, all statistical techniques presented in the text thus far are parametric techniques. **Parametric statistics** are *statistical techniques based on assumptions about the population from which the sample data are selected.* For example, if a t statistic is being used to conduct a hypothesis test about a population mean, the assumption is that the data being analyzed are randomly selected from a *normally* distributed population. The name *parametric statistics* refers to the fact that an assumption (here, normally distributed data) is being made about the data used to test or estimate the parameter (in this case, the population mean). In addition, the use of parametric statistics requires quantitative measurements that yield interval- or ratio-level data.

For data that do not meet the assumptions made about the population, or when the level of data being measured is qualitative, statistical techniques called nonparametric, or distribution-free, techniques are used. **Nonparametric statistics** *are based on fewer assumptions about the population and the parameters than are parametric statistics.* Sometimes they are referred to as *distribution-free* statistics because many of them can be used regardless of the shape of the population distribution. A variety of nonparametric statistics are available for use with nominal or ordinal data. Some require at least ordinal-level data, but others can be specifically targeted for use with nominal-level data.

Nonparametric techniques have the following advantages.

1. Sometimes there is no parametric alternative to the use of nonparametric statistics.
2. Certain nonparametric tests can be used to analyze nominal data.
3. Certain nonparametric tests can be used to analyze ordinal data.
4. The computations on nonparametric statistics are usually less complicated than those for parametric statistics, particularly for small samples.
5. Probability statements obtained from most nonparametric tests are exact probabilities.

Using nonparametric statistics also has some disadvantages.

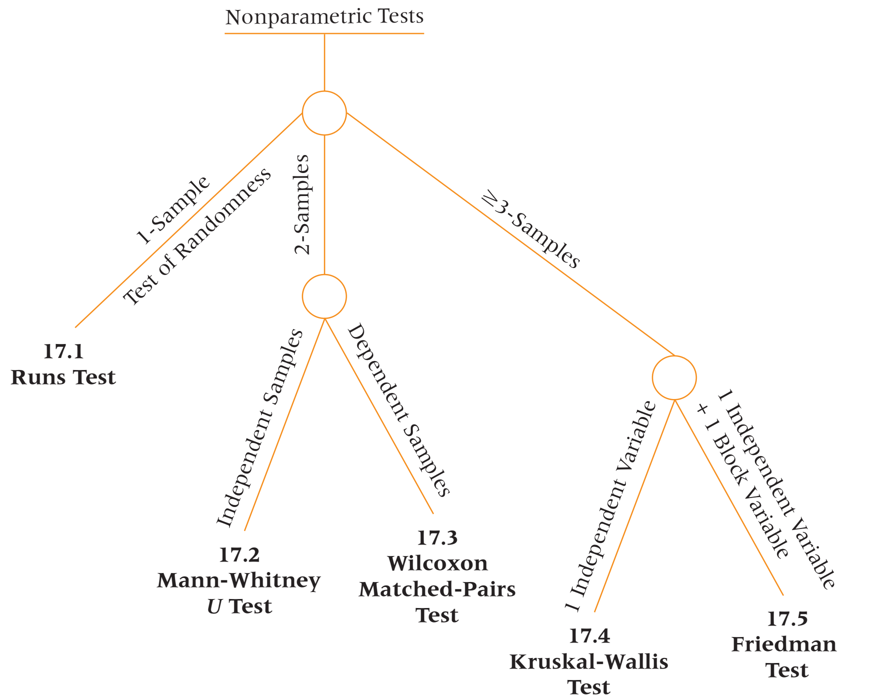
1. Nonparametric tests can be wasteful of data if parametric tests are available for use with the data.
2. Nonparametric tests are usually not as widely available and well known as parametric tests.
3. For large samples, the calculations for many nonparametric statistics can be tedious.

Entire courses and texts are dedicated to the study of nonparametric statistics. This text presents only some of the more important techniques: runs test, Mann-Whitney U test, Wilcoxon matched-pairs signed ranks test, Kruskal-Wallis test, Friedman test, Spearman's rank correlation coefficient, chi-square test of goodness-of-fit, and chi-square test of independence. The chi-square goodness-of-fit test and the chi-square test of independence were presented in Chapter 16. The others are presented in this chapter.

Figure 17.1 contains a tree diagram that displays all of the nonparametric techniques presented in this chapter, with the exception of Spearman's Rank Correlation, which is used to analyze the degree of association of two variables. As you peruse the tree diagram, you will see that there is a test of randomness, the runs test, two tests of the differences of two populations, the Mann-Whitney U test and the Wilcoxon matched-pairs signed rank test, and two tests of the differences of three or more populations—the Kruskal-Wallis test and the Friedman test.

FIGURE 17.1

Tree Diagram Taxonomy of Nonparametric Inferential Techniques



17.1 RUNS TEST

The one-sample **runs test** is *a nonparametric test of randomness*. The runs test is *used to determine whether the order or sequence of observations in a sample is random*. The runs test examines the number of "runs" of each of two possible characteristics that sample items may have. A *run* is a succession of observations that have a particular one of the characteristics. For example, if a sample of people contains both men and women, one run could be a continuous succession of women. In tossing coins, the outcome of three heads in a row would constitute a run, as would a succession of seven tails.

Suppose a researcher takes a random sample of 15 people who arrive at a Wal-Mart to shop. Eight of the people are women and seven are men. If these people arrive randomly at the store, it makes sense that the sequence of arrivals would have some mix of men and women, but not probably a perfect mix. That is, it seems unlikely (although possible) that the sequence of a random sample of such shoppers would be first eight women and then seven men. In such a case, there are two runs. Suppose, however, the sequence of shoppers is woman, man, woman, man, woman, and so on all the way through the sample. This would result in 15 "runs." Each of these cases is possible, but neither is highly likely in a random scenario. In fact, if there are just two runs, it seems possible that a group of women came shopping together followed by a group of men who did likewise. In that case, the observations would not be random. Similarly, a pattern of woman-man all the way through may make the business researcher suspicious that what has been observed is not really individual random arrivals, but actually random arrivals of couples.

In a random sample, the number of runs is likely to be somewhere between these extremes. What number of runs is reasonable? The one-sample runs test takes into consideration the size of the sample, n, the number observations in the sample having each characteristic, n_1, n_2 (man, woman, etc.), and the number of runs in the sample, R, to reach conclusions about hypotheses of randomness. The following hypotheses are tested by the one-sample runs test.

H_0: The observations in the sample are randomly generated.
H_a: The observations in the sample are not randomly generated.

The one-sample runs test is conducted differently for small samples than it is for large samples. Each test is presented here. First, we consider the small-sample case.

Small-Sample Runs Test

If both n_1 and n_2 are less than or equal to 20, the small-sample runs test is appropriate. In the example of shoppers with $n_1 = 7$ men and $n_2 = 8$ women, the small-sample runs test could be used to test for randomness. The test is carried out by comparing the observed number of runs, R, to critical values of runs for the given values of n_1 and n_2. The critical values of R are given in Tables A.11 and A.12 in the appendix for $\alpha = .05$. Table A.11 contains critical values of R for the lower tail of the distribution in which so few runs occur that the probability of that many runs or fewer runs occurring is less than .025 ($\alpha/2$). Table A.12 contains critical values of R for the upper tail of the distribution in which so many runs occur that the probability of that many runs or more occurring is less than .025 ($\alpha/2$). Any observed value of R that is less than or equal to the critical value of the lower tail (Table A.11) results in the rejection of the null hypothesis and the conclusion that the sample data are not random. Any observed value of R that is equal to or greater than the critical value in the upper tail (Table A.12) also results in the rejection of the null hypothesis and the conclusion that the sample data are not random.

As an example, suppose 26 cola drinkers are sampled randomly to determine whether they prefer regular cola or diet cola. The random sample contains 18 regular cola drinkers and eight diet cola drinkers. Let C denote regular cola drinkers and D denote diet cola drinkers. Suppose the sequence of sampled cola drinkers is DCCCCCDCCDCCCCDCDCCCDDDCCC. Is this sequence of cola drinkers evidence that the sample is not random? Applying the HTAB system of hypothesis testing to this problem results in:

HYPOTHESIZE:

STEP 1. The hypotheses tested follow.

H_0: The observations in the sample were generated randomly.
H_a: The observations in the sample were not generated randomly.

TEST:

STEP 2. Let n_1 denote the number of regular cola drinkers and n_2 denote the number of diet cola drinkers. Because $n_1 = 18$ and $n_2 = 8$, the small-sample runs test is the appropriate test.

STEP 3. Alpha is .05.

STEP 4. With $n_1 = 18$ and $n_2 = 8$, Table A.11 yields a critical value of 7 and Table A.12 yields a critical value of 17. If there are seven or fewer runs or 17 or more runs, the decision rule is to reject the null hypothesis.

STEP 5. The sample data are given as

DCCCCCDCCDCCCCDCDCCCDDDCCC

STEP 6. Tally the number of runs in this sample.

1	2	3	4	5	6	7	8	9	10	11	12
D	CCCCC	D	CC	D	CCCC	D	C	D	CCC	DDD	CCC

The number of runs, R, is 12.

ACTION:

STEP 7. Because the value of R falls between the critical values of 7 and 17, the decision is to not reject the null hypothesis. Not enough evidence is provided to declare that the data are not random.

BUSINESS IMPLICATION:

STEP 8. The cola researcher can proceed with the study under the assumption that the sample represents randomly selected cola drinkers.

FIGURE 17.2
Minitab Output for the Cola Example

```
Runs Test: Cola
Runs test for Cola
Runs above and below K = 1.69231

The observed number of runs = 12
The expected number of runs = 12.0769
18 observations above K, 8 below
* N is small, so the following approximation may be invalid.
P-value = 0.971
```

Minitab has the capability of analyzing data by using the runs test. Figure 17.2 is the Minitab output for the cola example runs test. Notice that the output includes the number of runs, 12, and the significance level of the test. For this analysis, diet cola was coded as a 1 and regular cola coded as a 2. The Minitab runs test is a two-tailed test, and the reported significance of the test is equivalent to a *p*-value. Because the significance is .9710, the decision is to not reject the null hypothesis.

Large-Sample Runs Test

Tables A.11 and A.12 do not contain critical values for n_1 and n_2 greater than 20. Fortunately, the sampling distribution of *R* is approximately normal with a mean and standard deviation of

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

The test statistic is a *z* statistic computed as

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{R - \left(\dfrac{2n_1n_2}{n_1 + n_2} + 1\right)}{\sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}}$$

The following hypotheses are being tested.

H_0: The observations in the sample were generated randomly.
H_a: The observations in the sample were not generated randomly.

The critical *z* values are obtained in the usual way by using alpha and Table A.5.

Consider the following manufacturing example. A machine produces parts that are occasionally flawed. When the machine is working in adjustment, flaws still occur but seem to happen randomly. A quality-control person randomly selects 50 of the parts produced by the machine today and examines them one at a time in the order that they were made. The result is 40 parts with no flaws and 10 parts with flaws. The sequence of no flaws (denoted by *N*) and flaws (denoted by *F*) is shown below. Using an alpha of .05, the quality controller tests to determine whether the machine is producing randomly (the flaws are occurring randomly).

NNN F NNNNNNN F NN FF NNNNNN F NNNN F NNNNNN
FFFF NNNNNNNNNNNN

HYPOTHESIZE:

STEP 1. The hypotheses follow.

H_0: The observations in the sample were generated randomly.
H_a: The observations in the sample were not generated randomly.

TEST:

STEP 2. The appropriate statistical test is the large-sample runs test. The test statistic is

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{R - \left(\dfrac{2n_1n_2}{n_1 + n_2} + 1\right)}{\sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}}$$

STEP 3. The value of alpha is .05.

STEP 4. This test is two tailed. Too few or too many runs could indicate that the machine is not producing flaws randomly. With $\alpha = .05$ and $\alpha/2 = .025$, the critical values are $z_{.025} = \pm 1.96$. The decision rule is to reject the null hypothesis if the observed value of the test statistic is greater than 1.96 or less than −1.96.

STEP 5. The preceding sequence provides the sample data. The value of n_1 is 40 and the value of n_2 is 10. The number of runs (R) is 13.

STEP 6.

$$\mu_R = \frac{2(40)(10)}{40 + 10} + 1 = 17$$

$$\sigma_R = \sqrt{\frac{2(40)(10)[2(40)(10) - 40 - 10]}{(40 + 10)^2(40 + 10 - 1)}} = 2.213$$

$$z = \frac{13 - 17}{2.213} = -1.81$$

ACTION:

STEP 7. Because the observed value of the test statistic, $z = -1.81$, is greater than the lower-tail critical value, $z = -1.96$, the decision is to not reject the null hypothesis.

BUSINESS IMPLICATION:

STEP 8. There is no evidence that the machine is not producing flaws randomly. If the null hypothesis had been rejected, there might be concern that the machine is producing flaws systematically and thereby is in need of inspection or repair.

Figure 17.3 is the Minitab output for this example. The value of K is the average of the observations. The data were entered into Minitab with a nonflaw coded as a 0 and a flaw as a 1. The value $K = .20$ is merely the average of these coded values. In Minitab, a run is a sequence of observations above or below this mean, which effectively yields the same thing as the number of 0s in a row (nonflaws) or number of 1s in a row (flaws). The nonflaws and flaws could have been coded as any two different numbers and the same results would have been achieved. The output shows the number of runs as 13 (the same number obtained manually) and a test significance (p-value) equal to .071. The test statistic is not significant at $\alpha = .05$ because the p-value is greater than .05.

FIGURE 17.3

Minitab Output for the Flawed Parts Example

```
Runs Test: Flaws
Runs Test for Flaws
Runs above and below K = 0.2

The observed number of runs = 13
The expected number of runs = 17
10 Observations above K, 40 below
* N is small, so the following approximation may be invalid.
P-value = 0.071
```

17.1 PROBLEMS

17.1 Test the following sequence of observations by using the runs test and $\alpha = .05$ to determine whether the process produced random results.
X X X Y X X Y Y Y X Y X Y X X Y Y Y Y X

17.2 Test the following sequence of observations by using the runs test and $\alpha = .05$ to determine whether the process produced random results.
M M N N N N N M M M M M M N N M M M M M N M M
N N N N N N N N N N N N N M M M M M M M M M M M

17.3 A process produced good parts and defective parts. A sample of 60 parts was taken and inspected. Eight defective parts were found. The sequence of good and defective parts was analyzed by using Minitab. The output is given here. With a two-tailed test and $\alpha = .05$, what conclusions can be reached about the randomness of the sample?

```
Runs Test: Defects
Runs test for Defects
Runs above and below K = 0.1333

The observed number of runs = 11
The expected number of runs = 14.8667
8 Observations above K, 52 below
P-value = 0.0264
```

17.4 A Watson Wyatt Worldwide survey showed that 58% of all Hispanic Americans are satisfied with their salary. Suppose a researcher randomly samples 27 Hispanic American workers and asks whether they are satisfied with their salary with the result that 15 say yes. The sequence of Yes and No responses is recorded and tested for randomness by means of Minitab. The output follows. Using an alpha of .05 and a two-tailed test, what could you conclude about the randomness of the sample?

```
Runs Test: Yes/No
Runs test for Yes/No
Runs above and below K = 0.5556

The observed number of runs = 18
The expected number of runs = 14.3333
15 Observations above K, 12 below
P-value = 0.1452
```

17.5 A Virginia Slims Opinion Poll by Roper Starch found that more than 70% of the women interviewed believe they have had more opportunity to succeed than their parents. Suppose a researcher in your state conducts a similar poll and asks the same question with the result that of 64 women interviewed, 40 believe they have had more opportunity to succeed than their parents. The sequence of responses to this question is given below with Y denoting Yes and N denoting No. Use the runs test and $\alpha = .05$ to test this sequence and determine whether the responses are random.

Y Y N Y Y N N Y Y Y N N Y N N Y Y Y Y Y N Y Y Y Y N N Y Y N N N Y Y Y N N Y
Y Y Y N Y N Y Y Y N N N Y N N Y Y Y Y Y N N Y Y Y Y

17.6 A survey conducted by the Ethics Resource Center discovered that 35% of all workers say that coworkers have committed some kind of office theft. Suppose a survey is conducted in your large company to ask the same question of 13 randomly selected employees. The results are that five of the sample say coworkers have committed some kind of office theft and eight say they are not aware of such infractions. The sequence of responses follows. (Y denotes a Yes answer and N denotes a No answer.) Use $\alpha = .05$ to test to determine whether this sequence represents a random sample.

N N N N Y Y Y N N N N Y Y

17.2 MANN-WHITNEY *U* TEST

Video

The **Mann-Whitney *U* test** is a *nonparametric counterpart of the t test used to compare the means of two independent populations.* This test was developed by Henry B. Mann and D. R. Whitney in 1947. Recall that the *t* test for independent samples presented in Chapter 10 can be used when data are at least interval in measurement and the populations are normally distributed. However, if the assumption of a normally distributed population is invalid or if the data are only ordinal in measurement, the *t* test should not be used. In such cases, the Mann-Whitney *U* test is an acceptable option for analyzing the data. The following assumptions underlie the use of the Mann-Whitney *U* test.

1. The samples are independent.
2. The level of data is at least ordinal.

The two-tailed hypotheses being tested with the Mann-Whitney *U* test are as follows.

H_0: The two populations are identical.
H_a: The two populations are not identical.

Computation of the *U* test begins by arbitrarily designating two samples as group 1 and group 2. The data from the two groups are combined into one group, with each data value retaining a group identifier of its original group. The pooled values are then ranked from 1 to n, with the smallest value being assigned a rank of 1. The sum of the ranks of values from group 1 is computed and designated as W_1 and the sum of the ranks of values from group 2 is designated as W_2.

The Mann-Whitney *U* test is implemented differently for small samples than for large samples. If both $n_1, n_2 \leq 10$, the samples are considered small. If either n_1 or n_2 is greater than 10, the samples are considered large.

Small-Sample Case

With small samples, the next step is to calculate a *U* statistic for W_1 and for W_2 as

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W_1 \quad \text{and} \quad U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - W_2$$

The test statistic is the smallest of these two *U* values. Both values do not need to be calculated; instead, one value of *U* can be calculated and the other can be found by using the transformation

$$U' = n_1 \cdot n_2 - U$$

Table A.13 contains p-values for U. To determine the p-value for a U from the table, let n_1 denote the size of the smaller sample and n_2 the size of the larger sample. Using the particular table in Table A.13 for n_1, n_2, locate the value of U in the left column. At the intersection of the U and n_1 is the p-value for a one-tailed test. For a two-tailed test, double the p-value shown in the table.

DEMONSTRATION PROBLEM 17.1

Is there a difference between health service workers and educational service workers in the amount of compensation employers pay them per hour? Suppose a random sample of seven health service workers is taken along with a random sample of eight educational service workers from different parts of the country. Each of their employers is interviewed and figures are obtained on the amount paid per hour for employee compensation for these workers. The data on the following page indicate total compensation per hour. Use a Mann-Whitney *U* test to determine whether these two populations are different in employee compensation.

Health Service Worker	Educational Service Worker
$20.10	$26.19
19.80	23.88
22.36	25.50
18.75	21.64
21.90	24.85
22.96	25.30
20.75	24.12
	23.45

Solution

HYPOTHESIZE:

STEP 1. The hypotheses are as follows.

H_0: The health service population is identical to the educational service population on employee compensation.

H_a: The health service population is not identical to the educational service population on employee compensation.

TEST:

STEP 2. Because we cannot be certain the populations are normally distributed, we chose a nonparametric alternative to the *t* test for independent populations: the small-sample Mann-Whitney *U* test.

STEP 3. Let alpha be .05.

STEP 4. If the final *p*-value from Table A.13 (after doubling for a two-tailed test here) is less than .05, the decision is to reject the null hypothesis.

STEP 5. The sample data were already provided.

STEP 6. We combine scores from the two groups and rank them from smallest to largest while retaining group identifier information.

Total Employee Compensation	Rank	Group
$18.75	1	H
19.80	2	H
20.10	3	H
20.75	4	H
21.64	5	E
21.90	6	H
22.36	7	H
22.96	8	H
23.45	9	E
23.88	10	E
24.12	11	E
24.85	12	E
25.30	13	E
25.50	14	E
26.19	15	E

$W_1 = 1 + 2 + 3 + 4 + 6 + 7 + 8 = 31$

$W_2 = 5 + 9 + 10 + 11 + 12 + 13 + 14 + 15 = 89$

$$U_1 = (7)(8) + \frac{(7)(8)}{2} - 31 = 53$$

$$U_2 = (7)(8) + \frac{(8)(9)}{2} - 89 = 3$$

Because U_2 is the smaller value of U, we use $U = 3$ as the test statistic for Table A.13. Because it is the smallest size, let $n_1 = 7$; $n_2 = 8$.

ACTION:

STEP 7. Table A.13 yields a *p*-value of .0011. Because this test is two tailed, we double the table *p*-value, producing a final *p*-value of .0022. Because the *p*-value is less than $\alpha = .05$, the null hypothesis is rejected. The statistical conclusion is that the populations are not identical.

BUSINESS IMPLICATIONS:

STEP 8. An examination of the total compensation figures from the samples indicates that employers pay educational service workers more per hour than they pay health service workers.

As shown in Figure 17.4, Minitab has the capability of computing a Mann-Whitney *U* test. The output includes a *p*-value of .0046 for the two-tailed test for Demonstration Problem 17.1. The decision based on the computer output is to reject the null hypothesis, which is consistent with what we computed. The difference in *p*-values is due to rounding error in the table.

Large-Sample Case

For large sample sizes, the value of *U* is approximately normally distributed. Using an average expected *U* value for groups of this size and a standard deviation of *U*'s allows computation of a *z* score for the *U* value. The probability of yielding a *z* score of this magnitude, given no difference between the groups, is computed. A decision is then made whether to reject the null hypothesis. A *z* score can be calculated from *U* by the following formulas.

LARGE-SAMPLE FORMULAS MANN-WHITNEY *U* TEST (17.1)

$$\mu_U = \frac{n_1 \cdot n_2}{2}, \quad \sigma_U = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}}, \quad z = \frac{U - \mu_U}{\sigma_U}$$

For example, the Mann-Whitney *U* test can be used to determine whether there is a difference in the average income of families who view PBS television and families who do not view PBS television. Suppose a sample of 14 families that have identified themselves as PBS television viewers and a sample of 13 families that have identified themselves as non-PBS television viewers are selected randomly.

HYPOTHESIZE:

STEP 1. The hypotheses for this example are as follows.

H_0: The incomes of PBS and non-PBS viewers are identical.
H_a: The incomes of PBS and non-PBS viewers are not identical.

TEST:

STEP 2. Use the Mann-Whitney *U* test for large samples.
STEP 3. Let $\alpha = .05$.

FIGURE 17.4
Minitab Output for Demonstration Problem 17.1

```
Mann-Whitney Test and CI: HS Worker, EdS Worker

              N    Median
HS Worker     7    20.750
EdS Worker    8    24.485
Point estimate for ETA1-ETA2 is -3.385
95.7 Percent CI for ETA1-ETA2 is (-5.370, -1.551)
W = 31.0
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is significant at 0.0046
```

TABLE 17.1

Income of PBS & Non-PBS Viewers

PBS	Non-PBS
$24,500	$41,000
39,400	32,500
36,800	33,000
43,000	21,000
57,960	40,500
32,000	32,400
61,000	16,000
34,000	21,500
43,500	39,500
55,000	27,600
39,000	43,500
62,500	51,900
61,400	27,800
53,000	
$n_1 = 14$	$n_2 = 13$

STEP 4. Because this test is two-tailed with $\alpha = .05$, the critical values are $z_{.025} = \pm 1.96$. If the test statistic is greater than 1.96 or less than -1.96, the decision is to reject the null hypothesis.

STEP 5. The average annual reported income for each family in the two samples is given in Table 17.1.

STEP 6. The first step toward computing a Mann-Whitney *U* test is to combine these two columns of data into one group and rank the data from lowest to highest, while maintaining the identification of each original group. Table 17.2 shows the results of this step.

Note that in the case of a tie, the ranks associated with the tie are averaged across the values that tie. For example, two incomes of $43,500 appear in the sample. These incomes represent ranks 19 and 20. Each value therefore is awarded a ranking of 19.5, or the average of 19 and 20.

If PBS viewers are designated as group 1, W_1 can be computed by summing the ranks of all the incomes of PBS viewers in the sample.

$$W_1 = 4 + 7 + 11 + 12 + 13 + 14 + 18 + 19.5 + 22 + 23 + 24 + 25 + 26 + 27 = 245.5$$

Then W_1 is used to compute the *U* value. Because $n_1 = 14$ and $n_2 = 13$, then

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W_1 = (14)(13) + \frac{(14)(15)}{2} - 245.5 = 41.5$$

Because $n_1, n_2 > 10$, *U* is approximately normally distributed, with a mean of

$$\mu_U = \frac{n_1 \cdot n_2}{2} = \frac{(14)(13)}{2} = 91$$

and a standard deviation of

$$\sigma_U = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(14)(13)(28)}{12}} = 20.6$$

A *z* value now can be computed to determine the probability of the sample *U* value coming from the distribution with $\mu_U = 91$ and $\sigma_U = 20.6$ if there is no difference in the populations.

$$z = \frac{U - \mu_U}{\sigma_U} = \frac{41.5 - 91}{20.6} = \frac{-49.5}{20.6} = -2.40$$

ACTION:

STEP 7. The observed value of *z* is -2.40, which is less than $Z_{\alpha/2} = -1.96$ so the results are in the rejection region. That is, there is a difference between the income of a PBS viewer and that of a non-PBS viewer. Examination of the sample data confirms that in general, the income of a PBS viewer is higher than that of a non-PBS viewer.

TABLE 17.2

Ranks of Incomes from Combined Groups of PBS and Non-PBS Viewers

Income	Rank	Group	Income	Rank	Group
$16,000	1	Non-PBS	39,500	15	Non-PBS
21,000	2	Non-PBS	40,500	16	Non-PBS
21,500	3	Non-PBS	41,000	17	Non-PBS
24,500	4	PBS	43,000	18	PBS
27,600	5	Non-PBS	43,500	19.5	PBS
27,800	6	Non-PBS	43,500	19.5	Non-PBS
32,000	7	PBS	51,900	21	Non-PBS
32,400	8	Non-PBS	53,000	22	PBS
32,500	9	Non-PBS	55,000	23	PBS
33,000	10	Non-PBS	57,960	24	PBS
34,000	11	PBS	61,000	25	PBS
36,800	12	PBS	61,400	26	PBS
39,000	13	PBS	62,500	27	PBS
39,400	14	PBS			

FIGURE 17.5

Minitab Output for the PBS Viewer Example

```
Mann-Whitney Test and CI: PBS, Non-PBS

               N    Median
PBS           14     43250
Non-PBS       13     32500
Point estimate for ETA1-ETA2 is 12500
95.1 Percent CI for ETA1-ETA2 is (3000,22000)
W = 245.5
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is significant at 0.0174
The test is significant at 0.0174 (adjusted for ties)
```

BUSINESS IMPLICATIONS:

STEP 8. The fact that PBS viewers have higher average income can affect the type of programming on PBS in terms of both trying to please present viewers and offering programs that might attract viewers of other income levels. In addition, fund-raising drives can be made to appeal to the viewers with higher incomes.

Assignment of PBS viewers to group 1 was arbitrary. If non-PBS viewers had been designated as group 1, the results would have been the same but the observed z value would have been positive.

Figure 17.5 is the Minitab output for this example. Note that Minitab does not produce a z value but rather yields the value of W and the probability of the test results occurring by chance (.0174). Because the p-value (.0174) is less than $\alpha = .05$, the decision based on the computer output is to reject the null hypothesis. The p-value of the observed test statistic ($z = -2.40$) is .0164. The difference is likely to be due to rounding error.

DEMONSTRATION PROBLEM 17.2

Do construction workers who purchase lunch from street vendors spend less per meal than construction workers who go to restaurants for lunch? To test this question, a researcher selects two random samples of construction workers, one group that purchases lunch from street vendors and one group that purchases lunch from restaurants. Workers are asked to record how much they spend on lunch that day. The data follow. Use the data and a Mann-Whitney U test to analyze the data to determine whether street-vendor lunches are significantly cheaper than restaurant lunches. Let $\alpha = .01$.

Vendor	Restaurant
\$2.75	\$4.10
3.29	4.75
4.53	3.95
3.61	3.50
3.10	4.25
4.29	4.98
2.25	5.75
2.97	4.10
4.01	2.70
3.68	3.65
3.15	5.11
2.97	4.80
4.05	6.25
3.60	3.89
	4.80
	5.50
$n_1 = 14$	$n_2 = 16$

Solution

HYPOTHESIZE:

STEP 1. The hypotheses follow.

H_0: The populations of construction-worker spending for lunch at vendors and restaurants are the same.

H_a: The population of construction-worker spending at vendors is shifted to the left of the population of construction-worker spending at restaurants.

TEST:

STEP 2. The large-sample Mann-Whitney *U* test is appropriate. The test statistic is the *z*.

STEP 3. Alpha is .01.

STEP 4. If the *p*-value of the sample statistic is less than .01, the decision is to reject the null hypothesis.

STEP 5. The sample data are given.

STEP 6. Determine the value of W_1 by combining the groups, while retaining group identification and ranking all the values from 1 to 30 (14 + 16), with 1 representing the smallest value.

Value	Rank	Group	Value	Rank	Group
$2.25	1	V	$4.01	16	V
2.70	2	R	4.05	17	V
2.75	3	V	4.10	18.5	R
2.97	4.5	V	4.10	18.5	R
2.97	4.5	V	4.25	20	R
3.10	6	V	4.29	21	V
3.15	7	V	4.53	22	V
3.29	8	V	4.75	23	R
3.50	9	R	4.80	24.5	R
3.60	10	V	4.80	24.5	R
3.61	11	V	4.98	26	R
3.65	12	R	5.11	27	R
3.68	13	V	5.50	28	R
3.89	14	R	5.75	29	R
3.95	15	R	6.25	30	R

Summing the ranks for the vendor sample gives

$$W_1 = 1 + 3 + 4.5 + 4.5 + 6 + 7 + 8 + 10 + 11 + 13 + 16 + 17 + 21 + 22 = 144$$

Solving for U, μ_U, and σ_U yields

$$U = (14)(16) + \frac{(14)(15)}{2} - 144 = 185 \qquad \mu_U = \frac{(14)(16)}{2} = 112$$

$$\sigma_U = \sqrt{\frac{(14)(16)(31)}{12}} = 24.1$$

Solving for the observed *z* value gives

$$z = \frac{185 - 112}{24.1} = 3.03$$

ACTION:

STEP 7. The *p*-value associated with $z = 3.03$ is .0012. The null hypothesis is rejected.

BUSINESS IMPLICATIONS:

STEP 8. The business researcher concludes that construction-worker spending at vendors is less than the spending at restaurants for lunches.

17.2 PROBLEMS

17.7 Use the Mann-Whitney *U* test and the following data to determine whether there is a significant difference between the values of group 1 and group 2. Let $\alpha = .05$.

Group 1	Group 2
15	23
17	14
26	24
11	13
18	22
21	23
13	18
29	21

17.8 The data shown represent two random samples gathered from two populations. Is there sufficient evidence in the data to determine whether the values of population 1 are significantly larger than the values of population 2? Use the Mann-Whitney *U* test and $\alpha = .01$.

Sample 1	Sample 2
224	203
256	218
231	229
222	230
248	211
283	230
241	209
217	223
240	219
255	236
216	227
	208
	214

17.9 Results of a survey by the National Center for Health Statistics indicated that people between 65 and 74 years of age contact a physician an average of 9.8 times per year. People 75 and older contact doctors an average of 12.9 times per year. Suppose you want to validate these results by taking your own samples. The following data represent the number of annual contacts people make with a physician. The samples are independent. Use a Mann-Whitney *U* test to determine whether the number of contacts with physicians by people 75 and older is greater than the number by people 65 to 74 years old. Let $\alpha = .01$.

65 to 74	75 and Older
12	16
13	15
8	10
11	17
9	13
6	12
11	14
	9
	13

17.10 Suppose 12 urban households and 12 rural households are selected randomly and each family is asked to report the amount spent on food at home annually. The results follow. Use a Mann-Whitney *U* test to determine whether there is a significant difference between urban and rural households in the amounts spent for food at home. Use $\alpha = .05$.

Urban	Rural	Urban	Rural
$2,110	$2,050	$1,950	$2,770
2,655	2,800	2,480	3,100
2,710	2,975	2,630	2,685
2,540	2,075	2,750	2,790
2,200	2,490	2,850	2,995
2,175	2,585	2,850	2,995

17.11 Does the male stock market investor earn significantly more than the female stock market investor? One study by the New York Stock Exchange showed that the male investor has an income of $46,400 and that the female investor has an income of $39,400. Suppose an analyst wanted to "prove" that the male investor earns more than the female investor. The following data represent random samples of male and female investors from across the United States. The analyst uses the Mann-Whitney *U* test to determine whether the male investor earns significantly more than the female investor for $\alpha = .01$. What does the analyst find?

Male	Female
$50,100	$41,200
47,800	36,600
45,000	44,500
51,500	47,800
55,000	42,500
53,850	47,500
51,500	40,500
63,900	28,900
57,800	48,000
61,100	42,300
51,000	40,000
	31,400

17.12 The National Association of Realtors reports that the median price of an existing single-family home in Denver, Colorado, is $225,100 and the median price of an existing single-family home in Hartford, Connecticut, is $233,700. Suppose a survey of 13 randomly selected single-family homes is taken in Denver and a survey of 15 randomly selected single-family homes is taken in Hartford with the resulting prices shown here. Use a Mann-Whitney *U* test to determine whether there is a significant difference in the price of a single-family home in these two cities. Let $\alpha = .05$.

Denver	Hartford
$234,157	$243,947
238,057	234,127
235,062	235,238
237,016	237,359
235,940	240,031
236,981	239,114
240,479	242,012
240,102	244,500
239,638	236,419
241,861	237,867
241,408	237,741
232,405	234,514
241,730	242,136
	236,333
	243,968

17.3 WILCOXON MATCHED-PAIRS SIGNED RANK TEST

Video

The Mann-Whitney *U* test presented in Section 17.2 is a nonparametric alternative to the *t* test for two *independent* samples. If the two samples are *related*, the *U* test is not applicable. A test that does handle related data is the **Wilcoxon matched-pairs signed rank test**, which serves as *a nonparametric alternative to the t test for two related samples.* Developed by Frank Wilcoxon in 1945, the Wilcoxon test, like the *t* test for two related samples, is used to analyze several different types of studies when the data of one group are related to the data in the other group, including before-and-after studies, studies in which measures are taken on the same person or object under two different conditions, and studies of twins or other relatives.

The Wilcoxon test utilizes the differences of the scores of the two matched groups in a manner similar to that of the *t* test for two related samples. After the difference scores have been computed, the Wilcoxon test ranks all differences regardless of whether the difference is positive or negative. The values are ranked from smallest to largest, with a rank of 1 assigned to the smallest difference. If a difference is negative, the rank is given a negative sign. The sum of the positive ranks is tallied along with the sum of the negative ranks. Zero differences representing ties between scores from the two groups are ignored, and the value of n is reduced accordingly. When ties occur between ranks, the ranks are averaged over the values. The smallest sum of ranks (either + or −) is used in the analysis and is represented by T. The Wilcoxon matched-pairs signed rank test procedure for determining statistical significance differs with sample size. When the number of matched pairs, n, is greater than 15, the value of T is approximately normally distributed and a z score is computed to test the null hypothesis. When sample size is small, $n \le 15$, a different procedure is followed.

Two assumptions underlie the use of this technique.

1. The paired data are selected randomly.
2. The underlying distributions are symmetrical.

The following hypotheses are being tested.
For two-tailed tests:

$$H_0\text{: } M_d = 0 \quad H_a\text{: } M_d \ne 0$$

For one-tailed tests:

$$H_0\text{: } M_d = 0 \quad H_a\text{: } M_d > 0$$

or

$$H_0\text{: } M_d = 0 \quad H_a\text{: } M_d < 0$$

where M_d is the median.

Small-Sample Case ($n \le 15$)

When sample size is small, a critical value against which to compare T can be found in Table A.14 to determine whether the null hypothesis should be rejected. The critical value is located by using n and α. Critical values are given in the table for $\alpha = .05, .025, .01$, and .005 for two-tailed tests and $\alpha = .10, .05, .02$, and.01 for one-tailed tests. If the observed value of T is less than or equal to the critical value of T, the decision is to reject the null hypothesis.

As an example, consider the survey by American Demographics that estimated the average annual household spending on healthcare. The U.S. metropolitan average was $1,800. Suppose six families in Pittsburgh, Pennsylvania, are matched demographically with six families in Oakland, California, and their amounts of household spending on healthcare for last year are obtained. The data follow on the next page.

Family Pair	Pittsburgh	Oakland
1	$1,950	$1,760
2	1,840	1,870
3	2,015	1,810
4	1,580	1,660
5	1,790	1,340
6	1,925	1,765

A healthcare analyst uses $\alpha = .05$ to test to determine whether there is a significant difference in annual household healthcare spending between these two cities.

HYPOTHESIZE:

STEP 1. The following hypotheses are being tested.

$$H_0: M_d = 0$$
$$H_a: M_d \neq 0$$

TEST:

STEP 2. Because the sample size of pairs is six, the small-sample Wilcoxon matched-pairs signed ranks test is appropriate if the underlying distributions are assumed to be symmetrical.

STEP 3. Alpha is .05.

STEP 4. From Table A.14, if the observed value of T is less than or equal to 1, the decision is to reject the null hypothesis.

STEP 5. The sample data were listed earlier.

STEP 6.

Family Pair	Pittsburgh	Oakland	d	Rank
1	$1,950	$1,760	+190	+4
2	1,840	1,870	−30	−1
3	2,015	1,810	+205	+5
4	1,580	1,660	−80	−2
5	1,790	1,340	+450	+6
6	1,925	1,765	+160	+3

$$T = \text{minimum of } (T_+, T_-)$$
$$T_+ = 4 + 5 + 6 + 3 = 18$$
$$T_- = 1 + 2 = 3$$
$$T = \text{minimum of } (18, 3) = 3$$

ACTION:

STEP 7. Because $T = 3$ is greater than critical $T = 1$, the decision is not to reject the null hypothesis.

BUSINESS IMPLICATIONS:

STEP 8. Not enough evidence is provided to declare that Pittsburgh and Oakland differ in annual household spending on healthcare. This information may be useful to healthcare providers and employers in the two cities and particularly to businesses that either operate in both cities or are planning to move from one to the other. Rates can be established on the notion that healthcare costs are about the same in both cities. In addition, employees considering transfers from one city to the other can expect their annual healthcare costs to remain about the same.

Large-Sample Case ($n > 15$)

For large samples, the T statistic is approximately normally distributed and a z score can be used as the test statistic. Formula 17.2 contains the necessary formulas to complete this procedure.

WILCOXON MATCHED-PAIRS SIGNED RANK TEST (17.2)

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

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

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

where

n = number of pairs
T = total ranks for either + or − differences, whichever is less in magnitude

This technique can be applied to the airline industry, where an analyst might want to determine whether there is a difference in the cost per mile of airfares in the United States between 1979 and 2009 for various cities. The data in Table 17.3 represent the costs per mile of airline tickets for a sample of 17 cities for both 1979 and 2009.

HYPOTHESIZE:

STEP 1. The analyst states the hypotheses as follows.

$$H_0: M_d = 0$$
$$H_a: M_d \neq 0$$

TEST:

STEP 2. The analyst applies a Wilcoxon matched-pairs signed rank test to the data to test the difference in cents per mile for the two periods of time. She assumes the underlying distributions are symmetrical.

STEP 3. Use $\alpha = .05$.

TABLE 17.3
Airline Ticket Costs for Various Cities

City	1979	2009	d	Rank
1	20.3	22.8	−2.5	−8
2	19.5	12.7	+6.8	+17
3	18.6	14.1	+4.5	+13
4	20.9	16.1	+4.8	+15
5	19.9	25.2	−5.3	−16
6	18.6	20.2	−1.6	−4
7	19.6	14.9	+4.7	+14
8	23.2	21.3	+1.9	+6.5
9	21.8	18.7	+3.1	+10
10	20.3	20.9	−0.6	−1
11	19.2	22.6	−3.4	−11.5
12	19.5	16.9	+2.6	+9
13	18.7	20.6	−1.9	−6.5
14	17.7	18.5	−0.8	−2
15	21.6	23.4	−1.8	−5
16	22.4	21.3	+1.1	+3
17	20.8	17.4	+3.4	+11.5

STEP 4. Because this test is two-tailed, $\alpha/2 = .025$ and the critical values are $z = \pm 1.96$. If the observed value of the test statistic is greater than 1.96 or less than −1.96, the null hypothesis is rejected.

STEP 5. The sample data are given in Table 17.3.

STEP 6. The analyst begins the process by computing a difference score, *d*. Which year's data are subtracted from the other does not matter as long as consistency in direction is maintained. For the data in Table 17.3, the analyst subtracted the 2009 figures from the 1979 figures. The sign of the difference is left on the difference score. Next, she ranks the differences without regard to sign, but the sign is left on the rank as an identifier. Note the tie for ranks 6 and 7; each is given a rank of 6.5, the average of the two ranks. The same applies to ranks 11 and 12.

After the analyst ranks all difference values regardless of sign, she sums the positive ranks (T_1) and the negative ranks (T_2). She then determines the T value from these two sums as the smallest T_1 or T_2.

$$T = \text{minimum of}\,(T_+, T_-)$$
$$T_+ = 17 + 13 + 15 + 14 + 6.5 + 10 + 9 + 3 + 11.5 = 99$$
$$T_- = 8 + 16 + 4 + 1 + 11.5 + 6.5 + 2 + 5 = 54$$
$$T = \text{minimum of}\,(99, 54) = 54$$

The T value is normally distributed for large sample sizes, with a mean and standard deviation of

$$\mu_T = \frac{(n)(n+1)}{4} = \frac{(17)(18)}{4} = 76.5$$
$$\sigma_T = \sqrt{\frac{(n)(n+1)(2n+1)}{24}} = \sqrt{\frac{(17)(18)(35)}{24}} = 21.1$$

The observed z value is

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{54 - 76.5}{21.1} = -1.07$$

ACTION:

STEP 7. The critical z value for this two-tailed test is $z_{.025} = \pm 1.96$. The observed $z = -1.07$, so the analyst fails to reject the null hypothesis. There is no significant difference in the cost of airline tickets between 1979 and 2009.

BUSINESS IMPLICATIONS:

STEP 8. Promoters in the airline industry can use this type of information (the fact that ticket prices have not increased significantly in 30 years) to sell their product as a good buy. In addition, industry managers could use it as an argument for raising prices.

DEMONSTRATION PROBLEM 17.3

During the 1980s and 1990s, U.S. businesses increasingly emphasized quality control. One of the arguments in favor of quality-control programs is that quality control can increase productivity. Suppose a company implemented a quality-control program and has been operating under it for 2 years. The company's president wants to determine whether worker productivity significantly increased since installation of the program. Company records contain the figures for items produced per worker during a sample of production runs 2 years ago. Productivity figures on the same workers are gathered now and compared to the previous figures. The following data represent items produced per hour. The company's statistical analyst uses the Wilcoxon matched-pairs signed rank test to determine whether there is a significant increase in per worker production for $\alpha = .01$.

Worker	Before	After	Worker	Before	After
1	5	11	11	2	6
2	4	9	12	5	10
3	9	9	13	4	9
4	6	8	14	5	7
5	3	5	15	8	9
6	8	7	16	7	6
7	7	9	17	9	10
8	10	9	18	5	8
9	3	7	19	4	5
10	7	9	20	3	6

Solution

HYPOTHESIZE:

STEP 1. The hypotheses are as follows.

$$H_0: M_d = 0$$
$$H_a: M_d < 0$$

TEST:

STEP 2. The analyst applies a Wilcoxon matched-pairs signed rank test to the data to test the difference in productivity from before to after. He assumes the underlying distributions are symmetrical.

STEP 3. Use $\alpha = .01$.

STEP 4. This test is one tailed. The critical value is $z = -2.33$. If the observed value of the test statistic is less than -2.33, the null hypothesis is rejected.

STEP 5. The sample data are as already given.

STEP 6. The analyst computes the difference values, and, because zero differences are to be eliminated, deletes worker 3 from the study. This reduces n from 20 to 19. He then ranks the differences regardless of sign. The differences that are the same (ties) receive the average rank for those values. For example, the differences for workers 4, 5, 7, 10, and 14 are the same. The ranks for these five are 7, 8, 9, 10, and 11, so each worker receives the rank of 9, the average of these five ranks.

Worker	Before	After	*d*	Rank
1	5	11	−6	−19
2	4	9	−5	−17
3	9	9	0	delete
4	6	8	−2	−9
5	3	5	−2	−9
6	8	7	+1	+3.5
7	7	9	−2	−9
8	10	9	+1	+3.5
9	3	7	−4	−14.5
10	7	9	−2	−9
11	2	6	−4	−14.5
12	5	10	−5	−17
13	4	9	−5	−17
14	5	7	−2	−9
15	8	9	−1	−3.5
16	7	6	+1	+3.5
17	9	10	−1	−3.5
18	5	8	−3	−12.5
19	4	5	−1	−3.5
20	3	6	−3	−12.5

The analyst determines the values of T_+, T_-, and T to be

$$T_+ = 3.5 + 3.5 + 3.5 = 10.5$$
$$T_- = 19 + 17 + 9 + 9 + 9 + 14.5 + 9 + 14.5 + 17 + 17 + 9 + 3.5 + 3.5 + 12.5 + 3.5 + 12.5 = 179.5$$
$$T = \text{minimum of } (10.5, 179.5) = 10.5$$

The mean and standard deviation of T are

$$\mu_T = \frac{(n)(n+1)}{4} = \frac{(19)(20)}{4} = 95$$
$$\sigma_T = \sqrt{\frac{(n)(n+1)(2n+1)}{24}} = \sqrt{\frac{(19)(20)(39)}{24}} = 24.8$$

The observed z value is

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{10.5 - 95}{24.8} = -3.41$$

ACTION:

STEP 7. The observed z value (−3.41) is in the rejection region, so the analyst rejects the null hypothesis. The productivity is significantly greater after the implementation of quality control at this company.

BUSINESS IMPLICATIONS:

STEP 8. Managers, the quality team, and any consultants can point to the figures as validation of the efficacy of the quality program. Such results could be used to justify further activity in the area of quality.

Figure 17.6 is Minitab output for Demonstration Problem 17.3. Minitab does not produce a z test statistic for the Wilcoxon matched-pairs signed rank test. Instead, it calculates a Wilcoxon statistic that is equivalent to T. A p-value of .000 is produced for this T value. The p-value of the observed $z = -3.41$ determined in Demonstration Problem 17.3 is .0003.

FIGURE 17.6

Minitab Output for Demonstration Problem 17.3

```
Wilcoxon Signed Rank Test: difference
Test of median = 0.000000 versus median < 0.000000
                   N for   Wilcoxon              Estimated
                N   Test   Statistic         P      Median
difference 20      19          10.5    0.0000     -2.000
```

17.3 PROBLEMS

17.13 Use the Wilcoxon matched-pairs signed rank test to determine whether there is a significant difference between the two groups of related data given. Use $\alpha = .10$. Assume the underlying distributions are symmetrical.

1	2	1	2
212	179	220	223
234	184	218	217
219	213	234	208
199	167	212	215
194	189	219	187
206	200	196	198
234	212	178	189
225	221	213	201

17.14 Use the Wilcoxon matched-pairs signed rank test and $\alpha = .05$ to analyze the before-and-after measurements given. Assume the underlying distributions are symmetrical.

Before	After
49	43
41	29
47	30
39	38
53	40
51	43
51	46
49	40
38	42
54	50
46	47
50	47
44	39
49	49
45	47

17.15 A corporation owns a chain of several hundred gasoline stations on the eastern seaboard. The marketing director wants to test a proposed marketing campaign by running ads on some local television stations and determining whether gasoline sales at a sample of the company's stations increase after the advertising. The following data represent gasoline sales for a day before and a day after the advertising campaign. Use the Wilcoxon matched-pairs signed rank test to determine whether sales increased significantly after the advertising campaign. Let $\alpha = .05$. Assume the underlying distributions are symmetrical.

Station	Before	After
1	$10,500	$12,600
2	8,870	10,660
3	12,300	11,890
4	10,510	14,630
5	5,570	8,580
6	9,150	10,115
7	11,980	14,350
8	6,740	6,900
9	7,340	8,890
10	13,400	16,540
11	12,200	11,300
12	10,570	13,330
13	9,880	9,990
14	12,100	14,050
15	9,000	9,500
16	11,800	12,450
17	10,500	13,450

17.16 Most supermarkets across the United States have invested heavily in optical scanner systems to expedite customer checkout, increase checkout productivity, and improve product accountability. These systems are not 100% effective, and items often have to be scanned several times. Sometimes items are entered into the manual cash register because the scanner cannot read the item number. In general, do optical scanners register significantly more items than manual entry systems do? The following data are from an experiment in which a supermarket selected 14 of its best checkers and measured their productivity both when using a scanner and when working manually. The data show the number of items checked per hour by each method. Use a Wilcoxon matched-pairs signed rank test and $\alpha = .05$ to test the difference. Assume the underlying distributions are symmetrical.

Checker	Manual	Scanner
1	426	473
2	387	446
3	410	421
4	506	510
5	411	465
6	398	409
7	427	414
8	449	459
9	407	502
10	438	439
11	418	456
12	482	499
13	512	517
14	402	437

17.17 American attitudes toward big business change over time and probably are cyclical. Suppose the following data represent a survey of 20 American adults taken in 1990 and again in 2009 in which each adult was asked to rate American big business overall on a scale from 1 to 100 in terms of positive opinion. A response of 1 indicates a low opinion and a response of 100 indicates a high opinion. Use a Wilcoxon matched-pairs signed rank test to determine whether the scores from 2009 are significantly higher than the scores from 1990. Use $\alpha = .10$. Assume the underlying distributions are symmetrical.

Person	1990	2009
1	49	54
2	27	38
3	39	38
4	75	80
5	59	53
6	67	68
7	22	43
8	61	67
9	58	73
10	60	55
11	72	58
12	62	57
13	49	63
14	48	49
15	19	39
16	32	34
17	60	66
18	80	90
19	55	57
20	68	58

17.18 Suppose 16 people in various industries are contacted in 2008 and asked to rate business conditions on several factors. The ratings of each person are tallied into a "business optimism" score. The same people are contacted in 2009 and asked to do the same thing. The higher the score, the more optimistic the person is. Shown here are the 2008 and 2009 scores for the 16 people. Use a Wilcoxon matched-pairs signed rank test to determine whether people were less optimistic in 2009 than in 2008. Assume the underlying distributions are symmetrical and that alpha is .05.

Industry	April 2008	April 2009
1	63.1	57.4
2	67.1	66.4
3	65.5	61.8
4	68.0	65.3
5	66.6	63.5
6	65.7	66.4
7	69.2	64.9
8	67.0	65.2
9	65.2	65.1
10	60.7	62.2
11	63.4	60.3
12	59.2	57.4
13	62.9	58.2
14	69.4	65.3
15	67.3	67.2
16	66.8	64.1

17.4 KRUSKAL-WALLIS TEST

Video

The *nonparametric alternative to the one-way analysis of variance* is the **Kruskal-Wallis test**, developed in 1952 by William H. Kruskal and W. Allen Wallis. Like the one-way analysis of variance, the Kruskal-Wallis test is used to determine whether $c \geq 3$ samples come from the same or different populations. Whereas the one-way ANOVA is based on the assumptions of normally distributed populations, independent groups, at least interval level data, and equal population variances, the Kruskal-Wallis test can be used to analyze ordinal data and is not based on any assumption about population shape. The Kruskal-Wallis test is based on the assumption that the c groups are independent and that individual items are selected randomly.

The hypotheses tested by the Kruskal-Wallis test follow.

H_0: The c populations are identical.

H_a: At least one of the c populations is different.

This test determines whether all of the groups come from the same or equal populations or whether at least one group comes from a different population.

The process of computing a Kruskal-Wallis K statistic begins with ranking the data in all the groups together, as though they were from one group. The smallest value is awarded a 1. As usual, for ties, each value is given the average rank for those tied values. Unlike one-way ANOVA, in which the raw data are analyzed, the Kruskal-Wallis test analyzes the ranks of the data.

Formula 17.3 is used to compute a Kruskal-Wallis K statistic.

KRUSKAL-WALLIS TEST (17.3)

$$K = \frac{12}{n(n+1)}\left(\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right) - 3(n+1)$$

where

c = number of groups
n = total number of items
T_j = total of ranks in a group
n_j = number of items in a group
$K \approx \chi^2$, with df $= c - 1$

TABLE 17.4

Number of Office Patients per Doctor

Two Partners	Three or More Partners	HMO
13	24	26
15	16	22
20	19	31
18	22	27
23	25	28
	14	33
	17	

TABLE 17.5

Kruskal-Wallis Analysis of Physicians' Patients

Two Partners	Three or More Partners	HMO	
1	12	14	
3	4	9.5	
8	7	17	
6	9.5	15	
11	13	16	
	2	18	
	5		
$T_1 = 29$	$T_2 = 52.5$	$T_3 = 89.5$	
$n_1 = 5$	$n_2 = 7$	$n_3 = 6$	$n = 18$

$$\sum_{j=1}^{3} \frac{T_j^2}{n_j} = \frac{(29)^2}{5} + \frac{(52.5)^2}{7} + \frac{(89.5)^2}{6} = 1{,}897$$

The K value is approximately chi-square distributed, with $c - 1$ degrees of freedom as long as n_j is not less than 5 for any group.

Suppose a researcher wants to determine whether the number of physicians in an office produces significant differences in the number of office patients seen by each physician per day. She takes a random sample of physicians from practices in which (1) there are only two partners, (2) there are three or more partners, or (3) the office is a health maintenance organization (HMO). Table 17.4 shows the data she obtained.

Three groups are targeted in this study, so $c = 3$, and $n = 18$ physicians, with the numbers of patients ranked for these physicians. The researcher sums the ranks within each column to obtain T_j, as shown in Table 17.5.

The Kruskal-Wallis K is

$$K = \frac{12}{18(18+1)}(1{,}897) - 3(18+1) = 9.56$$

The critical chi-square value is $\chi^2_{\alpha,\text{df}}$. If $\alpha = .05$ and df for $c - 1 = 3 - 1 = 2$, $\chi^2_{.05,2} = 5.9915$. This test is always one-tailed, and the rejection region is always in the right tail of the distribution. Because $K = 9.56$ is larger than the critical χ^2 value, the researcher rejects the null hypothesis. The number of patients seen in the office by a physician is not the same in these three sizes of offices. Examination of the values in each group reveals that physicians in two-partner offices see fewer patients per physician in the office, and HMO physicians see more patients per physician in the office.

Figure 17.7 is the Minitab computer output for this example. The statistic H printed in the output is equivalent to the K statistic calculated here (both K and H are 9.56).

FIGURE 17.7

Minitab Output for the Physicians' Patients Example

```
Kruskal-Wallis Test on Patients

Group       N   Median   Ave Rank      Z
1           5    18.00        5.8  -1.82
2           7    19.00        7.5  -1.27
3           6    27.50       14.9   3.04
Overall    18                 9.5

H = 9.56  DF = 2  P = 0.008
H = 9.57  DF = 2  P = 0.008 (adjusted for ties)
```

DEMONSTRATION PROBLEM 17.4

Agribusiness researchers are interested in determining the conditions under which Christmas trees grow fastest. A random sample of equivalent-size seedlings is divided into four groups. The trees are all grown in the same field. One group is left to grow naturally, one group is given extra water, one group is given fertilizer spikes, and one group is given fertilizer spikes and extra water. At the end of one year, the seedlings are measured for growth (in height). These measurements are shown for each group. Use the Kruskal-Wallis test to determine whether there is a significant difference in the growth of trees in these groups. Use $\alpha = .01$.

Group 1 (native)	Group 2 (+water)	Group 3 (+fertilizer)	Group 4 (+water and fertilizer)
8 in.	10 in.	11 in.	18 in.
5	12	14	20
7	11	10	16
11	9	16	15
9	13	17	14
6	12	12	22

Solution

Here, $n = 24$, and $n_j = 6$ in each group.

HYPOTHESIZE:

STEP 1. The hypotheses follow.

H_0: group 1 = group 2 = group 3 = group 4

H_a: At least one group is different.

TEST:

STEP 2. The Kruskal-Wallis K is the appropriate test statistic.

STEP 3. Alpha is .01.

STEP 4. The degrees of freedom are $c - 1 = 4 - 1 = 3$. The critical value of chi-square is $\chi^2_{.01,3} = 11.3449$. If the observed value of K is greater than 11.3449, the decision is to reject the null hypothesis.

STEP 5. The data are as shown previously.

STEP 6. Ranking all group values yields the following.

1	2	3	4	
4	7.5	10	22	
1	13	16.5	23	
3	10	7.5	19.5	
10	5.5	19.5	18	
5.5	15	21	16.5	
2	13	13	24	
$T_1 = 25.5$	$T_2 = 64.0$	$T_3 = 87.5$	$T_4 = 123.0$	
$n_1 = 6$	$n_2 = 6$	$n_3 = 6$	$n_4 = 6$	$n = 24$

$$\sum_{j=1}^{c} \frac{T_j^2}{n_j} = \frac{(25.5)^2}{6} + \frac{(64)^2}{6} + \frac{(87.5)^2}{6} + \frac{(123)^2}{6} = 4{,}588.6$$

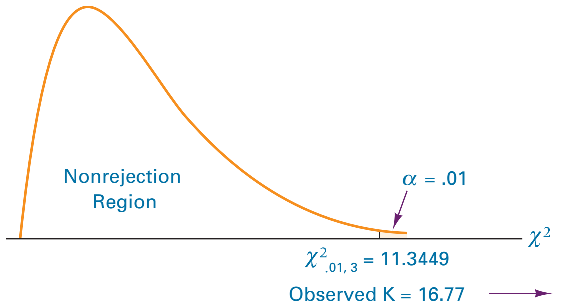
$$K = \frac{12}{24(24 + 1)}(4{,}588.6) - 3(24 + 1) = 16.77$$

ACTION:

STEP 7. The observed K value is 16.77 and the critical $\chi^2_{.01,3} = 11.3449$. Because the observed value is greater than the table value, the null hypothesis is rejected. There is a significant difference in the way the trees grow.

BUSINESS IMPLICATIONS:

STEP 8. From the increased heights in the original data, the trees with both water and fertilizer seem to be doing the best. However, these are sample data; without analyzing the pairs of samples with nonparametric multiple comparisons (not included in this text), it is difficult to conclude whether the water/fertilizer group is actually growing faster than the others. It appears that the trees under natural conditions are growing more slowly than the others. The following diagram shows the relationship of the observed *K* value and the critical chi-square value.



17.4 PROBLEMS

17.19 Use the Kruskal-Wallis test to determine whether groups 1 through 5 come from different populations. Let $\alpha = .01$.

1	2	3	4	5
157	165	219	286	197
188	197	257	243	215
175	204	243	259	235
174	214	231	250	217
201	183	217	279	240
203		203		233
				213

17.20 Use the Kruskal-Wallis test to determine whether there is a significant difference in the following groups. Use $\alpha = .05$.

Group 1	19	21	29	22	37	42	
Group 2	30	38	35	24	29		
Group 3	39	32	41	44	30	27	33

17.21 Is there a difference in the amount of customers' initial deposits when they open savings accounts according to geographic region of the United States? To test this question, an analyst selects savings and loan offices of equal size from four regions of the United States. The offices selected are located in areas having similar economic and population characteristics. The analyst randomly selects adult customers who are opening their first savings account and obtains the following dollar amounts. Use the Kruskal-Wallis test to determine whether there is a significant difference between geographic regions. Use $\alpha = .05$.

Region 1	Region 2	Region 3	Region 4
$1,200	$225	$675	$1,075
450	950	500	1,050
110	100	1,100	750
800	350	310	180
375	275	660	330
200			680
			425

STATISTICS IN BUSINESS TODAY

Heavy Internet Users: Who Are They?

The penetration rate of the Web in the United States was around 40% in the year 2000. Since less than half of the population was regularly using the Web, it was important for marketers to profile who were these people and determine their particular needs and wants. However, by the year 2005, the Web penetration rate was approaching 70%, thus closing the gap between this particular market segment and the U.S. population in general. Because of this, Henry Assael, a researcher at the Stern School of Business at New York University, conducted a study focusing on the demographic and psychographic profiles of heavy Internet users—those who use the Web for at least 20 hours a week instead of studying Web users in general. Although these heavy users probably represent only about 20% of the Web users, they account for possibly as much as 50% of all usage.

Based on a survey of 5,000 respondents, Assael discovered that heavy users are 40% more likely to be in the 18-to-34-year age bracket and 20% more likely to be in the 35-to-44-year age bracket than the general population of Web users. Although 12% more females than males use the Web in general, the gap is only 2% for heavy users. The percentage of Web users who have never been married is 60% higher for heavy users than for the general population of users. There is a higher percentage of heavy users in the $50,000 to $99,999 and the $100,000 to $150,000 income brackets than in the general population of Web users. Thirty-seven percent of heavy Web users have a bachelor's degree compared to 28% of all Web users. Heavy Web users tend to work more hours per week. The study showed that 39% of heavy Web users work more than 40 hours per week compared to only 27% of all Web users. Another study showed that the heaviest Internet users are also watching the most TV.

Nonparametric techniques can be used in studies similar to this one. For example, one study published in the year 2001 found that males average over 1,300 minutes per month online compared to about 1,200 for females. If the distributions of online usage are unknown, the Mann-Whitney *U* test could be used to test to determine if there is a significant difference in online usage between males and females in today's market. Furthermore, suppose the demographics of online users can be broken down by income levels such as: under $25,000, $25,000 to under $50,000, $50,000 to $99,999, and more than $100,000. A Kruskal-Wallis test could be used to determine if there is a significant difference in online usage by income level. In addition, a Wilcoxon matched-pairs signed rank test could be used to determine if online usage has significantly increased from one year to the next for the same set of users.

Source: Henry Assael, "A Demographic and Psychographic Profile of Heavy Internet Users and Users by Type of Internet Usage," *Journal of Advertising Research*, vol. 45, no. 1 (March, 2005), pp. 93–123. "Who Goes There?" *The Wall Street Journal* (October 29, 2001), p. R4. http://blog.nielsen.com/nielsenwire/online-mobile/heavy-internet-users-also-watch-more-tv/

17.22 Does the asking price of a new car vary according to whether the dealership is in a small town, a city, or a suburban area? To test this question, a researcher randomly selects dealerships selling Pontiacs in the state of Illinois. The researcher goes to these dealerships posing as a prospective buyer and makes a serious inquiry as to the asking price of a new Pontiac Grand Prix sedan (each having the same equipment). The following data represent the results of this sample. Is there a significant difference between prices according to the area in which the dealership is located? Use the Kruskal-Wallis test and $\alpha = .05$.

Small Town	City	Suburb
$21,800	$22,300	$22,000
22,500	21,900	22,600
21,750	21,900	22,800
22,200	22,650	22,050
21,600	21,800	21,250
		22,550

17.23 A survey by the U.S. Travel Data Center showed that a higher percentage of Americans travel to the ocean/beach for vacation than to any other destination. Much further behind in the survey, and virtually tied for second place, were the mountains and small/rural towns. How long do people stay at vacation destinations? Does the length of stay differ according to location? Suppose the following data were taken from a survey of vacationers who were asked how many nights they stay at a destination when on vacation. Use a Kruskal-Wallis test to determine whether there is a significant difference in the duration of stay by type of vacation destination. Let $\alpha = .05$.

Amusement Park	Lake Area	City	National Park
0	3	2	2
1	2	2	4
1	3	3	3
0	5	2	4
2	4	3	3
1	4	2	5
0	3	3	4
	5	3	4
	2	1	
		3	

17.24 Do workers on different shifts get different amounts of sleep per week? Some people believe that shift workers who regularly work the graveyard shift (12:00 A.M. to 8:00 A.M.) or swing shift (4:00 P.M. to 12:00 A.M.) are unable to get the same amount of sleep as day workers because of family schedules, noise, amount of daylight, and other factors. To test this theory, a researcher samples workers from day, swing, and graveyard shifts and asks each worker to keep a sleep journal for one week. The following data represent the number of hours of sleep per week per worker for the different shifts. Use the Kruskal-Wallis test to determine whether there is a significant difference in the number of hours of sleep per week for workers on these shifts. Use $\alpha = .05$.

Day Shift	Swing Shift	Graveyard Shift
52	45	41
57	48	46
53	44	39
56	51	49
55	48	42
50	54	35
51	49	52
	43	

17.5 FRIEDMAN TEST

Video

The **Friedman test**, developed by M. Friedman in 1937, is *a nonparametric alternative to the randomized block design* discussed in Chapter 11. The randomized block design has the same assumptions as other ANOVA procedures, including observations are drawn from normally distributed populations. When this assumption cannot be met or when the researcher has ranked data, the Friedman test provides a nonparametric alternative.

Three assumptions underlie the Friedman test.

1. The blocks are independent.
2. No interaction is present between blocks and treatments.
3. Observations within each block can be ranked.

The hypotheses being tested are as follows.

H_0: The treatment populations are equal.
H_a: At least one treatment population yields larger values than at least one other treatment population.

The first step in computing a Friedman test is to convert all raw data to ranks (unless the data are already ranked). However, unlike the Kruskal-Wallis test where all data are ranked together, the data in a Friedman test are ranked *within* each block from smallest (1) to largest (c). Each block contains c ranks, where c is the number of treatment levels. Using these ranks, the Friedman test will test to determine whether it is likely that the different treatment levels (columns) came from the same population. Formula 17.4 is used to calculate the test statistic, which is approximately chi-square distributed with df $= c - 1$ if $c > 4$ or when $c = 3$ and $b > 9$, or when $c = 4$ and $b > 4$.

FRIEDMAN TEST (17.4)

$$\chi_r^2 = \frac{12}{bc(c+1)}\sum_{j=1}^{c} R_j^2 - 3b(c+1)$$

where

c = number of treatment levels (columns)
b = number of blocks (rows)
R_j = total of ranks for a particular treatment level (column)
j = particular treatment level (column)
$\chi_r^2 \approx \chi^2$, with df $= c - 1$

As an example, suppose a manufacturing company assembles microcircuits that contain a plastic housing. Managers are concerned about an unacceptably high number of the products that sustained housing damage during shipment. The housing component is made by four different suppliers. Managers have decided to conduct a study of the plastic housing by randomly selecting five housings made by each of the four suppliers. To determine whether a supplier is consistent during the production week, one housing is selected for each day of the week. That is, for each supplier, a housing made on Monday is selected, one made on Tuesday is selected, and so on.

In analyzing the data, the treatment variable is supplier and the treatment levels are the four suppliers. The blocking effect is day of the week with each day representing a block level. The quality control team wants to determine whether there is any significant difference in the tensile strength of the plastic housing by supplier. The data are given here (in pounds per inch).

Day	Supplier 1	Supplier 2	Supplier 3	Supplier 4
Monday	62	63	57	61
Tuesday	63	61	59	65
Wednesday	61	62	56	63
Thursday	62	60	57	64
Friday	64	63	58	66

HYPOTHESIZE:

STEP 1. The hypotheses follow.

H_0: The supplier populations are equal.
H_a: At least one supplier population yields larger values than at least one other supplier population.

TEST:

STEP 2. The quality researchers do not feel they have enough evidence to conclude that the observations come from normally distributed populations. Because they are analyzing a randomized block design, the Friedman test is appropriate.

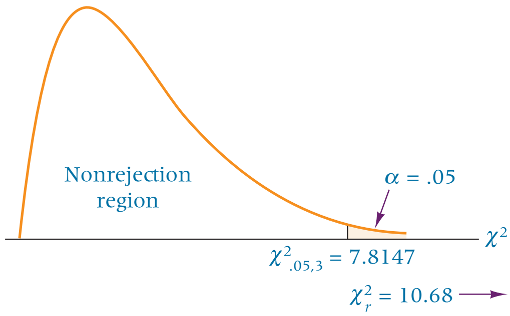
STEP 3. Let $\alpha = .05$.

STEP 4. For four treatment levels (suppliers), $c = 4$ and df $= 4 - 1 = 3$. The critical value is $\chi^2_{.05,3} = 7.8147$. If the observed chi-square is greater than 7.8147, the decision is to reject the null hypothesis.

STEP 5. The sample data are as given.

STEP 6. The calculations begin by ranking the observations in each row with 1 designating the rank of the smallest observation. The ranks are then summed for each column, producing R_j. The values of R_j are squared and then summed. Because the study is concerned with five days of the week, five blocking levels are used and $b = 5$. The value of R_j is computed as shown in the following table.

FIGURE 17.8
Distribution for Tensile Strength Example



Day	Supplier 1	Supplier 2	Supplier 3	Supplier 4
Monday	3	4	1	2
Tuesday	3	2	1	4
Wednesday	2	3	1	4
Thursday	3	2	1	4
Friday	3	2	1	4
R_j	14	13	5	18
R_j^2	196	169	25	324

$$\sum_{j=1}^{4} R_j^2 = (196 + 169 + 25 + 324) = 714$$

$$\chi_r^2 = \frac{12}{bc(c+1)}\sum_{j=1}^{c} R_j^2 - 3b(c+1) = \frac{12}{5(4)(4+1)}(714) - 3(5)(4+1) = 10.68$$

ACTION:

STEP 7. Because the observed value of $\chi_r^2 = 10.68$ is greater than the critical value, $\chi^2_{.05,3} = 7.8147$, the decision is to reject the null hypothesis.

BUSINESS IMPLICATIONS:

STEP 8. Statistically, there is a significant difference in the tensile strength of housings made by different suppliers. The sample data indicate that supplier 3 is producing housings with a lower tensile strength than those made by other suppliers and that supplier 4 is producing housings with higher tensile strength. Further study by managers and a quality team may result in attempts to bring supplier 3 up to standard on tensile strength or perhaps cancellation of the contract.

Figure 17.8 displays the chi-square distribution for df = 3 along with the critical value, the observed value of the test statistic, and the rejection region. Figure 17.9 is the Minitab output for the Friedman test. The computer output contains the value of χ_r^2 referred to as S along with the *p*-value of .014, which informs the researcher that the null hypothesis is rejected at an alpha of .05. Additional information is given about the medians and the column sum totals of ranks.

FIGURE 17.9
Minitab Output for the Tensile Strength Example

```
Friedman Test: Tensile Strength Versus Supplier Blocked by Day
S = 10.68   DF = 3   P = 0.014

                                 Sum
                                  of
Supplier   N   Est Median    Ranks
1          5       62.125     14.0
2          5       61.375     13.0
3          5       56.875      5.0
4          5       64.125     18.0

Grand median = 61.125
```

DEMONSTRATION PROBLEM 17.5

A market research company wants to determine brand preference for refrigerators. Five companies contracted with the research company to have their products be included in the study. As part of the study, the research company randomly selects 10 potential refrigerator buyers and shows them one of each of the five brands. Each survey participant is then asked to rank the refrigerator brands from 1 to 5. The results of these rankings are given in the table. Use the Friedman test and $\alpha = .01$ to determine whether there are any significant differences between the rankings of these brands.

Solution

HYPOTHESIZE:

STEP 1. The hypotheses are as follows.

H_0: The brand populations are equal.

H_a: At least one brand population yields larger values than at least one other brand population.

TEST:

STEP 2. The market researchers collected ranked data that are ordinal in level. The Friedman test is the appropriate test.

STEP 3. Let $\alpha = .01$.

STEP 4. Because the study uses five treatment levels (brands), $c = 5$ and df $= 5 - 1 = 4$. The critical value is $\chi^2_{.01,4} = 13.2767$. If the observed chi-square is greater than 13.2767, the decision is to reject the null hypothesis.

STEP 5. The sample data follow.

STEP 6. The ranks are totaled for each column, squared, and then summed across the column totals. The results are shown in the table.

Individual	Brand A	Brand B	Brand C	Brand D	Brand E
1	3	5	2	4	1
2	1	3	2	4	5
3	3	4	5	2	1
4	2	3	1	4	5
5	5	4	2	1	3
6	1	5	3	4	2
7	4	1	3	2	5
8	2	3	4	5	1
9	2	4	5	3	1
10	3	5	4	2	1
R_j	26	37	31	31	25
R_j^2	676	1,369	961	961	625
					$\Sigma R_j^2 = 4{,}592$

The value of χ_r^2 is

$$\chi_r^2 = \frac{12}{bc(c+1)}\sum_{j=1}^{c} R_j^2 - 3b(c+1) = \frac{12}{10(5)(5+1)}(4{,}592) - 3(10)(5+1) = 3.68$$

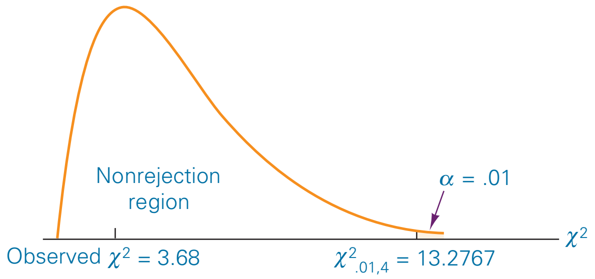
ACTION:

STEP 7. Because the observed value of $\chi_r^2 = 3.68$ is not greater than the critical value, $\chi^2_{.01,4} = 13.2767$, the researchers fail to reject the null hypothesis.

BUSINESS IMPLICATIONS:

STEP 8. Potential refrigerator purchasers appear to have no significant brand preference. Marketing managers for the various companies might want to develop strategies for positively distinguishing their product from the others.

The chi-square distribution for four degrees of freedom, produced by Minitab, is shown with the observed test statistic and the critical value. In addition, Minitab output for the Friedman test is shown. Note that the *p*-value is .451, which underscores the decision not to reject the null hypothesis at $\alpha = .01$.



Minitab Friedman Output:

```
Friedman Test: Rank Versus Brand Blocked by Individual

S = 3.68  DF = 4  P = 0.451

                        Sum
                Est      of
Brand    N   Median  Ranks
1       10    2.300   26.0
2       10    4.000   37.0
3       10    3.000   31.0
4       10    3.000   31.0
5       10    1.700   25.0

Grand median = 2.800
```

17.5 PROBLEMS

17.25 Use the following data to test to determine whether there are any differences between treatment levels. Let $\alpha = .05$.

		Treatment				
		1	2	3	4	5
Block	1	200	214	212	215	208
	2	198	211	214	217	206
	3	207	206	213	216	207
	4	213	210	215	219	204
	5	211	209	210	221	205

17.26 Use the Friedman test and $\alpha = .05$ to test the following data to determine whether there is a significant difference between treatment levels.

		Treatment					
		1	2	3	4	5	6
Block	1	29	32	31	38	35	33
	2	33	35	30	42	34	31
	3	26	34	32	39	36	35
	4	30	33	35	41	37	32
	5	33	31	32	35	37	36
	6	31	34	33	37	36	35
	7	26	32	35	43	36	34
	8	32	29	31	38	37	35
	9	30	31	34	41	39	35

17.27 An experiment is undertaken to study the effects of four different medical treatments on the recovery time for a medical disorder. Six physicians are involved in the study. One patient with the disorder is sampled for each physician under each treatment, resulting in 24 patients in the study. Recovery time in days is the observed measurement. The data are given here. Use the Friedman test and $\alpha = .01$ to determine whether there is a significant difference in recovery times for the four different medical treatments.

		Treatment			
		1	**2**	**3**	**4**
	1	3	7	5	4
	2	4	5	6	3
Physician	3	3	6	5	4
	4	3	6	7	4
	5	2	6	7	3
	6	4	5	7	3

17.28 Does the configuration of the workweek have any impact on productivity? This question is raised by a researcher who wants to compare the traditional 5-day workweek with a 4-day workweek and a workweek with three 12-hour days and one 4-hour day. The researcher conducts the experiment in a factory making small electronic parts. He selects 10 workers who spend a month working under each type of workweek configuration. The researcher randomly selects one day from each of the 3 months (three workweek configurations) for each of the 10 workers. The observed measurement is the number of parts produced per day by each worker. Use the Friedman test to determine whether there is a difference in productivity by workweek configuration.

		Workweek Configuration		
		5 Days	**4 Days**	**3½ Days**
	1	37	33	28
	2	44	38	36
	3	35	29	31
	4	41	40	36
Worker	5	38	39	35
	6	34	27	23
	7	43	38	39
	8	39	35	32
	9	41	38	37
	10	36	30	31

17.29 Shown here is Minitab output from a Friedman test. What is the size of the experimental design in terms of treatment levels and blocks? Discuss the outcome of the experiment in terms of any statistical conclusions.

```
Friedman Test: Observations Versus Treatment Blocked by Block

S = 2.04   DF = 3   P = 0.564

                               Sum
                        Est     of
Treatment    N    Median  Ranks
1            5     3.250   15.0
2            5     2.000   10.0
3            5     2.750   11.0
4            5     4.000   14.0

Grand median = 3.000
```

17.30 Shown here is Minitab output for a Friedman test. Discuss the experimental design and the outcome of the experiment.

```
Friedman Test: Observations Versus Treatment Blocked by Block

S = 13.71  DF = 4   P = 0.009

                             Sum
                        Est    of
Treatment   N   Median  Ranks
1           7   21.000   12.0
2           7   24.000   14.0
3           7   29.800   30.0
4           7   27.600   26.0
5           7   27.600   23.0

Grand median = 26.000
```

17.6 SPEARMAN'S RANK CORRELATION

In Chapter 12, the Pearson product-moment correlation coefficient, r, was presented and discussed as a technique to measure the amount or degree of association between two variables. The Pearson r requires at least interval level of measurement for the data. When only ordinal-level data or ranked data are available, **Spearman's rank correlation**, r_s, can be used to analyze the degree of association of two variables. Charles E. Spearman (1863–1945) developed this correlation coefficient.

The formula for calculating a Spearman's rank correlation is as follows:

SPEARMAN'S RANK CORRELATION (17.7)

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

where

n = number of pairs being correlated
d = the difference in the ranks of each pair

The Spearman's rank correlation formula is derived from the Pearson product-moment formula and utilizes the ranks of the n pairs instead of the raw data. The value of d is the difference in the ranks of each pair.

The process begins by the assignment of ranks within each group. The difference in ranks between each group (d) is calculated by subtracting the rank of a member of one group from the rank of its associated member of the other group. The differences (d) are then squared and summed. The number of pairs in the groups is represented by n.

The interpretation of r_s values is similar to the interpretation of r values. Positive correlations indicate that high values of one variable tend to be associated with high values of the other variable, and low values of one variable tend to be associated with low values of the other variable. Correlations near +1 indicate high positive correlations, and correlations near −1 indicate high negative correlations. Negative correlations indicate that high values of one variable tend to be associated with low values of the other variable, and vice versa. Correlations near zero indicate little or no association between variables.

Listed in Table 17.6 are the average prices in dollars per 100 pounds for choice spring lambs and choice heifers over a 10-year period. The data were published by the National Agricultural Statistics Service of the U.S. Department of Agriculture. Suppose we want to determine the strength of association of the prices between these two commodities by using Spearman's rank correlation.

TABLE 17.6

Spring Lamb and Choice Heifer Prices over a 10-Year Period

Year	Choice Spring Lamb Prices ($/100 lbs.)	Choice Heifer Prices ($/100 lbs.)
1	77.91	65.46
2	82.00	64.18
3	89.20	65.66
4	74.37	59.23
5	66.42	65.68
6	80.10	69.55
7	69.78	67.81
8	72.09	67.39
9	92.14	82.06
10	96.31	84.40

TABLE 17.7

Calculations of Spearman's Rank Correlation for Lamb and Heifer Prices over a 10-Year Period

Year	Rank: Lamb	Rank: Heifer	d	d^2
1	5	3	2	4
2	7	2	5	25
3	8	4	4	16
4	4	1	3	9
5	1	5	−4	16
6	6	8	−2	4
7	2	7	−5	25
8	3	6	−3	9
9	9	9	0	0
10	10	10	0	0
				$\Sigma d^2 = 108$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(108)}{10(10^2 - 1)} = 0.345$$

The lamb prices are ranked and the heifer prices are ranked. The difference in ranks is computed for each year. The differences are squared and summed, producing $\Sigma d^2 = 108$. The number of pairs, n, is 10. The value of $r_s = 0.345$ indicates that there is a very modest positive correlation between lamb and heifer prices. The calculations of this Spearman's rank correlation are enumerated in Table 17.7.

DEMONSTRATION PROBLEM 17.6

How strong is the correlation between crude oil prices and prices of gasoline at the pump? In an effort to estimate this association, an oil company analyst gathered the data shown over a period of several months. She lets crude oil prices be represented by the market value of a barrel of West Texas intermediate crude and gasoline prices be the estimated average price of regular unleaded gasoline in a certain city. She computes a Spearman's rank correlation for these data.

Crude Oil	Gasoline
$14.60	$3.25
10.50	3.26
12.30	3.28
15.10	3.26
18.35	3.32
22.60	3.44
28.90	3.56
31.40	3.60
26.75	3.54

Solution

Here, $n = 9$. When the analyst ranks the values within each group and computes the values of d and d^2, she obtains the following.

Crude Oil	Gasoline	d	d^2
3	1	+2	4
1	2.5	−1.5	2.25
2	4	−2	4
4	2.5	+1.5	2.25
5	5	0	0
6	6	0	0
8	8	0	0
9	9	0	0
7	7	0	0
			$\Sigma d^2 = 12.5$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(12.5)}{9(9^2 - 1)} = +.896$$

A high positive correlation is computed between the price of a barrel of West Texas intermediate crude and a gallon of regular unleaded gasoline.

17.6 PROBLEMS

17.31 Compute a Spearman's rank correlation for the following variables to determine the degree of association between the two variables.

x	y
23	201
41	259
37	234
29	240
25	231
17	209
33	229
41	246
40	248
28	227
19	200

17.32 The following data are the ranks for values of the two variables, x and y. Compute a Spearman's rank correlation to determine the degree of relation between the two variables.

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

17.33 Compute a Spearman's rank correlation for the following data.

x	y	x	y
99	108	80	124
67	139	57	162
82	117	49	145
46	168	91	102

17.34 Over a period of a few months, is there a strong correlation between the value of the U.S. dollar and the prime interest rate? The following data represent a sample of these quantities over a period of time. Compute a Spearman's rank correlation to determine the strength of the relationship between prime interest rates and the value of the dollar.

Dollar Value	Prime Rate	Dollar Value	Prime Rate
92	9.3	88	8.4
96	9.0	84	8.1
91	8.5	81	7.9
89	8.0	83	7.2
91	8.3		

17.35 Shown here are the percentages of consumer loans with payments that are 30 days or more overdue for both bank credit cards and home equity loans over a 14-year period according to the American Bankers Association. Compute a Spearman's rank correlation to determine the degree of association between these two variables.

Year	Bank Credit Card	Home Equity Loan
1	2.51%	2.07%
2	2.86	1.95
3	2.33	1.66
4	2.54	1.77
5	2.54	1.51
6	2.18	1.47
7	3.34	1.75
8	2.86	1.73
9	2.74	1.48
10	2.54	1.51
11	3.18	1.25
12	3.53	1.44
13	3.51	1.38
14	3.11	1.30

17.36 Shown here are the net tonnage figures for total pig iron and raw steel output in the United States as reported by the American Iron and Steel Institute over a 12-year period. Use these data to calculate a Spearman's rank correlation to determine the degree of association between production of pig iron and raw steel over this period. Was the association strong? Comment.

Year	Total Pig Iron (net tons)	Raw Steel (net tons)
1	43,952,000	81,606,000
2	48,410,000	89,151,000
3	55,745,000	99,924,000
4	55,873,000	97,943,000
5	54,750,000	98,906,000
6	48,637,000	87,896,000
7	52,224,000	92,949,000
8	53,082,000	97,877,000
9	54,426,000	100,579,000
10	56,097,000	104,930,000
11	54,485,000	105,309,478
12	54,679,000	108,561,182

17.37 Is there a correlation between the number of companies listed on the New York Stock Exchange in a given year and the number of equity issues on the American Stock Exchange? Shown on the next page are the values for these two variables over an 11-year period. Compute a Spearman's rank correlation to determine the degree of association between these two variables.

Year	Number of Companies on NYSE	Number of Equity Issues on AMEX
1	1774	1063
2	1885	1055
3	2088	943
4	2361	1005
5	2570	981
6	2675	936
7	2907	896
8	3047	893
9	3114	862
10	3025	769
11	2862	765

How Is the Doughnut Business?

Decision Dilemma Solved

The Dunkin' Donuts researchers' dilemma is that in each of the three studies presented, the assumptions underlying the use of parametric statistics are in question or have not been met. The distribution of the data is unknown bringing into question the normal distribution assumption or the level of data is only ordinal. For each study, a nonparametric technique presented in this chapter could be appropriately used to analyze the data.

The differences in doughnut sizes according to machine can be analyzed using the Kruskal-Wallis test. The independent variable is machine with four levels of classification. The dependent variable is size of doughnut in centimeters. The Kruskal-Wallis test is not based on any assumption about population shape. The following Minitab output is from a Kruskal-Wallis test on the machine data presented in the Decision Dilemma.

Kruskal-Wallis Test: Diameter Versus Machine

```
Kruskal-Wallis Test on Diameter

Machine     N   Median   Ave Rank      Z
1           5   7.520        10.4  -0.60
2           6   7.415         3.5  -3.57
3           5   7.550        12.6   0.22
4           7   7.700        20.0   3.74
Overall    23                12.0

H = 19.48 DF = 3  P = 0.000

H = 19.51 DF = 3  P = 0.000 (adjusted for ties)
```

Because the H statistic (Minitab's equivalent to the K statistic) has a p-value of .000, there is a significant difference in the diameter of the doughnut according to machine at $\alpha = .001$. An examination of median values reveals that machine 4 is producing the largest doughnuts and machine 2 the smallest.

How well did the advertising work? One way to address this question is to perform a before-and-after test of the number of doughnuts sold. The nonparametric alternative to the matched-pairs t test is the Wilcoxon matched-pairs signed rank test. The analysis for these data is:

Before	After	d	Rank
301	374	−73	−9
198	187	11	4
278	332	−54	−7
205	212	−7	−3
249	243	6	2
410	478	−68	−8
360	386	−26	−6
124	141	−17	−5
253	251	2	1
190	264	−74	−10

$T_+ = 4 + 2 + 1 = 7$

$T_- = 9 + 7 + 3 + 8 + 6 + 5 + 10 = 48$

observed $T = \min(T_+, T_-) = 7$

critical T for .025 and $n = 10$ is 8

Using a two-sided test and $\alpha = .05$, the critical T value is 8. Because the observed T is 7, the decision is to reject the null hypothesis. There is a significant difference between the before and after number of donuts sold. An observation of the ranks and raw data reveals that a majority of the stores experienced an increase in sales after the advertising campaign.

Do bigger stores have greater sales? Because the data are given as ranks, it is appropriate to use Spearman's Rank Correlation to determine the extent of the correlation between these two variables. Shown on the next page are the calculations of a Spearman's Rank correlation for this problem.

Sales	Size	d	d^2
6	7	−1	1
2	2	0	0
3	6	−3	9
7	5	2	4
5	4	1	1
1	1	0	0
4	3	1	1
			$\Sigma d^2 = 16$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6(16)}{7(49 - 1)} = .714$$

There is a relatively strong correlation (.714) between sales and size of store. It is not, however, a perfect correlation, which leaves room for other factors that may determine a store's sales such as location, attractiveness of store, population density, number of employees, management style, and others.

ETHICAL CONSIDERATIONS

The researcher should be aware of all assumptions underlying the usage of statistical techniques. Many parametric techniques have level-of-data requirements and assumptions about the distribution of the population or assumptions about the parameters. Inasmuch as these assumptions and requirements are not met, the researcher sets herself or himself up for misuse of statistical analysis. Spurious results can follow, and misguided conclusions can be reached. Nonparametric statistics can be used in many cases to avoid such pitfalls. In addition, some nonparametric statistics require at least ordinal-level data.

SUMMARY

Nonparametric statistics are a group of techniques that can be used for statistical analysis when the data are less than interval in measurement or when assumptions about population parameters, such as shape of the distribution, cannot be met. Nonparametric tests offer several advantages. Sometimes the nonparametric test is the only technique available, with no parametric alternative. Nonparametric tests can be used to analyze nominal- or ordinal-level data. Computations from nonparametric tests are usually simpler than those used with parametric tests. Probability statements obtained from most nonparametric tests are exact probabilities. Nonparametric techniques also have some disadvantages. They are wasteful of data whenever a parametric technique can be used. Nonparametric tests are not as widely available as parametric tests. For large sample sizes, the calculations of nonparametric statistics can be tedious.

Many of the parametric techniques presented in this text have corresponding nonparametric techniques. The six nonparametric statistical techniques presented here are the runs test, the Mann-Whitney U test, the Wilcoxon matched-pairs signed rank test, the Kruskal-Wallis test, the Friedman test, and Spearman's rank correlation.

The runs test is a nonparametric test of randomness. It is used to determine whether the order of sequence of observations in a sample is random. A run is a succession of observations that have a particular characteristic. If data are truly random, neither a very high number of runs nor a very small number of runs is likely to be present.

The Mann-Whitney U test is a nonparametric version of the t test of the means from two independent samples. When the assumption of normally distributed data cannot be met or if the data are only ordinal in level of measurement, the Mann-Whitney U test can be used in place of the t test. The Mann-Whitney U test—like many nonparametric tests—works with the ranks of data rather than the raw data.

The Wilcoxon matched-pairs signed rank test is used as an alternative to the t test for related measures when assumptions cannot be met or if the data are ordinal in measurement. In contrast to the Mann-Whitney U test, the Wilcoxon test is used when the data are related in some way. The Wilcoxon test is used to analyze the data by ranks of the differences of the raw data.

The Kruskal-Wallis test is a nonparametric one-way analysis of variance technique. It is particularly useful when the assumptions underlying the F test of the parametric one-way ANOVA cannot be met. The Kruskal-Wallis test is usually used when the researcher wants to determine whether three or more groups or samples are from the same or equivalent populations. This test is based on the assumption that the sample items are selected randomly and that the groups are independent. The raw data are converted to ranks and the Kruskal-Wallis test is used to analyze the ranks with the equivalent of a chi-square statistic.

The Friedman test is a nonparametric alternative to the randomized block design. Friedman's test is computed by ranking the observations within each block and then summing the ranks for each treatment level. The resulting test statistic χ^2 is approximately chi-square distributed.

If two variables contain data that are ordinal in level of measurement, a Spearman's rank correlation can be used to determine the amount of relationship or association between the variables. Spearman's rank correlation coefficient is a nonparametric alternative to Pearson's product-moment correlation coefficient. Spearman's rank correlation coefficient is interpreted in a manner similar to the Pearson r.

KEY TERMS

Flash Cards

- Friedman test
- Kruskal-Wallis test
- Mann-Whitney U test
- nonparametric statistics
- parametric statistics
- runs test
- Spearman's rank correlation
- Wilcoxon matched-pairs signed rank test

FORMULAS

Large-sample runs test

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1$$

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

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{R - \left(\dfrac{2n_1n_2}{n_1 + n_2} + 1\right)}{\sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}}$$

Mann-Whitney U test

Small sample:

$$U_1 = n_1n_2 + \frac{n_1(n_1 + 1)}{2} - W_1$$

$$U_2 = n_1n_2 + \frac{n_2(n_2 + 1)}{2} - W_2$$

$$U' = n_1 \cdot n_2 - U$$

Large sample:

$$\mu_U = \frac{n_1 \cdot n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}}$$

$$z = \frac{U - \mu_U}{\sigma_U}$$

Wilcoxon matched-pair signed rank test

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

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

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

Kruskal-Wallis test

$$K = \frac{12}{n(n + 1)}\left(\sum_{j=1}^{c} \frac{T_j^2}{n_j}\right) - 3(n + 1)$$

Friedman test

$$\chi_r^2 = \frac{12}{bc(c + 1)}\sum_{j=1}^{c} R_j^2 - 3b(c + 1)$$

Spearman's rank correlation

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

SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

17.38 Use the runs test to determine whether the sample is random. Let alpha be .05.

1 1 1 1 2 2 2 2 2 2 2 1 1 1 2 2 2
2 2 2 2 2 1 2 1 2 2 1 1 1 1 2 2 2

17.39 Use the Mann-Whitney U test and $\alpha = .01$ to determine whether there is a significant difference between the populations represented by the two samples given here.

Sample 1	Sample 2
573	547
532	566
544	551
565	538
540	557
548	560
536	557
523	547

17.40 Use the Wilcoxon matched-pairs signed rank test to determine whether there is a significant difference between the related populations represented by the matched pairs given here. Assume $\alpha = .05$.

Group 1	Group 2
5.6	6.4
1.3	1.5
4.7	4.6
3.8	4.3
2.4	2.1
5.5	6.0
5.1	5.2
4.6	4.5
3.7	4.5

17.41 Use the Kruskal-Wallis test and $\alpha = .01$ to determine whether the four groups come from different populations.

Group 1	Group 2	Group 3	Group 4
6	4	3	1
11	13	7	4
8	6	7	5
10	8	5	6
13	12	10	9
7	9	8	6
10	8	5	7

17.42 Use the Friedman test to determine whether the treatment groups come from different populations. Let alpha be .05.

Block	Group 1	Group 2	Group 3	Group 4
1	16	14	15	17
2	8	6	5	9
3	19	17	13	18
4	24	26	25	21
5	13	10	9	11
6	19	11	18	13
7	21	16	14	15

17.43 Compute a Spearman's rank correlation to determine the degree of association between the two variables.

Variable 1	Variable 2
101	87
129	89
133	84
147	79
156	70
179	64
183	67
190	71

TESTING YOUR UNDERSTANDING

17.44 Commercial fish raising is a growing industry in the United States. What makes fish raised commercially grow faster and larger? Suppose that a fish industry study is conducted over the three summer months in an effort to determine whether the amount of water allotted per fish makes any difference in the speed with which the fish grow. The following data represent the inches of growth of marked catfish in fish farms for different volumes of water per fish. Use $\alpha = .01$ to test whether there is a significant difference in fish growth by volume of allotted water.

1 Gallon per Fish	5 Gallons per Fish	10 Gallons per Fish
1.1	2.9	3.1
1.4	2.5	2.4
1.7	2.6	3.0
1.3	2.2	2.3
1.9	2.1	2.9
1.4	2.0	1.9
2.1	2.7	

17.45 Manchester Partners International claims that 60% of the banking executives who lose their job stay in banking, whereas 40% leave banking. Suppose 40 people who have lost their job as a banking executive are contacted and are asked whether they are still in banking. The results follow. Test to determine whether this sample appears to be random on the basis of the sequence of those who have left banking and those who have not. Let L denote left banking and S denote stayed in banking. Let $\alpha = .05$.

S S L S L L S S S S L S S L L L S S L
L L L S S L S S S S S S L L S L S S L S

17.46 Three machines produce the same part. Ten different machine operators work these machines. A quality team wants to determine whether the machines are producing parts that are significantly different from each other in weight. The team devises an experimental design in which a random part is selected from each of the 10 machine operators on each machine. The results follow.

Using alpha of .05, test to determine whether there is a difference in machines.

Operator	Machine 1	Machine 2	Machine 3
1	231	229	234
2	233	232	231
3	229	233	230
4	232	235	231
5	235	228	232
6	234	237	231
7	236	233	230
8	230	229	227
9	228	230	229
10	237	238	234

17.47 In some firefighting organizations, you must serve as a firefighter for some period of time before you can become part of the emergency medical service arm of the organization. Does that mean EMS workers are older, on average, than traditional firefighters? Use the data shown and $\alpha = .05$ to test whether EMS workers are significantly older than firefighters. Assume the two groups are independent and you do no want to use a *t* test to analyze the data.

Firefighters	EMS Workers	Firefighters	EMS Workers
23	27	32	39
37	29	24	33
28	30	21	30
25	33	27	28
41	28		27
36	36		30

17.48 Automobile dealers usually advertise in the yellow pages of the telephone book. Sometimes they have to pay to be listed in the white pages, and some dealerships opt to save money by omitting that listing, assuming most people will use the yellow pages to find the telephone number. A 2-year study is conducted with 20 car dealerships where in one year the dealer is listed in the white pages and the other year it is not. Ten of the dealerships are listed in the white pages the first year and the other 10 are listed there in the second year in an attempt to control for economic cycles. The following data represent the numbers of units sold per year. Is there a significant difference between the number of units sold when the dealership is listed in the white pages and the number sold when it is not listed? Assume all companies are continuously listed in the yellow pages, that the *t* test is not appropriate, and that $\alpha = .01$.

Dealer	With Listing	Without Listing
1	1,180	1,209
2	874	902
3	1,071	862
4	668	503
5	889	974
6	724	675
7	880	821
8	482	567
9	796	602
10	1,207	1,097
11	968	962
12	1,027	1,045
13	1,158	896
14	670	708
15	849	642
16	559	327
17	449	483
18	992	978
19	1,046	973
20	852	841

17.49 Suppose you want to take a random sample of GMAT test scores to determine whether there is any significant difference between the GMAT scores for the test given in March and the scores for the test given in June. You gather the following data from a sample of persons who took each test. Use the Mann-Whitney *U* test to determine whether there is a significant difference in the two test results. Let $\alpha = .10$.

March	June
540	350
570	470
600	630
430	590
500	610
510	520
530	460
560	550
550	530
490	570

17.50 Does impulse buying really increase sales? A market researcher is curious to find out whether the location of packages of chewing gum in a grocery store really has anything to do with volume of gum sales. As a test, gum is moved to a different location in the store every Monday for 4 weeks (four locations). To control the experiment for type of gum, six different brands are moved around. Sales representatives keep track of how many packs of each type of gum are sold every Monday for the 4 weeks. The results follow. Test to determine whether there are any differences in the volume of gum sold at the various locations. Let $\alpha = .05$.

		Location			
		1	2	3	4
Brand	A	176	58	111	120
	B	156	62	98	117
	C	203	89	117	105
	D	183	73	118	113
	E	147	46	101	114
	F	190	83	113	115

17.51 Does deodorant sell better in a box or without additional packaging? An experiment in a large store is designed in which, for one month, all deodorants are sold packaged in a box and, during a second month, all deodorants are removed from the box and sold without packaging. Is there a significant difference in the number of units of deodorant sold with and without the additional packaging? Let $\alpha = .05$.

Deodorant	Box	No Box
1	185	170
2	109	112
3	92	90
4	105	87
5	60	51
6	45	49
7	25	11
8	58	40
9	161	165
10	108	82
11	89	94
12	123	139
13	34	21
14	68	55
15	59	60
16	78	52

17.52 Some people drink coffee to relieve stress on the job. Is there a correlation between the number of cups of coffee consumed on the job and perceived job stress? Suppose the data shown represent the number of cups of coffee consumed per week and a stress rating for the job on a scale of 0 to 100 for nine managers in the same industry. Determine the correlation between these two variables, assuming you do not want to use the Pearson product-moment correlation coefficient.

Cups of Coffee per Week	Job Stress
25	80
41	85
16	35
0	45
11	30
28	50
34	65
18	40
5	20

17.53 A Gallup/Air Transport Association survey showed that in a recent year, 52% of all air trips were for pleasure/personal and 48% were for business. Suppose the organization randomly samples 30 air travelers and asks them to state the purpose of their trip. The results are shown here with B denoting business and P denoting personal. Test the sequence of these data to determine whether the data are random. Let $\alpha = .05$.

B P B P B B P B P P B P B P P
P B P B B P B P P B B P P B B

17.54 Does a statistics course improve a student's mathematics skills, as measured by a national test? Suppose a random sample of 13 students takes the same national mathematics examination just prior to enrolling in a statistics course and just after completing the course. Listed are the students' quantitative scores from both examinations. Use $\alpha = .01$ to determine whether the scores after the statistics course are significantly higher than the scores before.

Student	Before	After
1	430	465
2	485	475
3	520	535
4	360	410
5	440	425
6	500	505
7	425	450
8	470	480
9	515	520
10	430	430
11	450	460
12	495	500
13	540	530

17.55 Should male managers wear a tie during the workday to command respect and demonstrate professionalism? Suppose a measurement scale has been developed that generates a management professionalism score. A random sample of managers in a high-tech industry is selected for the study, some of whom wear ties at work and others of whom do not. One subordinate is selected randomly from each manager's department and asked to complete the scale on their boss's professionalism. Analyze the data taken from these independent groups to determine whether the managers with the ties received significantly higher professionalism scores. Let $\alpha = .05$.

With Tie	Without Tie
27	22
23	16
25	25
22	19
25	21
26	24
21	20
25	19
26	23
28	26
22	17

17.56 Many fast-food restaurants have soft drink dispensers with preset amounts, so that when the operator merely pushes a button for the desired drink the cup is automatically filled. This method apparently saves time and seems to increase worker productivity. To test this conclusion, a researcher randomly selects 18 workers from the fast-food industry, 9 from a restaurant with automatic soft drink dispensers and 9 from a comparable restaurant with manual soft drink dispensers. The samples are independent. During a comparable hour, the amount of sales rung up by the worker is recorded. Assume that $\alpha = .01$ and that a t test is not appropriate. Test whether workers with automatic dispensers are significantly more productive (higher sales per hour).

Automatic Dispenser	Manual Dispenser
$153	$105
128	118
143	129
110	114
152	125
168	117
144	106
137	92
118	126

17.57 A particular metal part can be produced at different temperatures. All other variables being equal, a company would like to determine whether the strength of the metal part is significantly different for different temperatures. Given are the strengths of random samples of parts produced under different temperatures. Use $\alpha = .01$ and determine whether there is a significant difference in the strength of the part for different temperatures.

45°	55°	70°	85°
216	228	219	218
215	224	220	216
218	225	221	217
216	222	223	221
219	226	224	218
214	225		217

17.58 Is there a strong correlation between the number of miles driven by a salesperson and sales volume achieved? Data were gathered from nine salespeople who worked territories of similar size and potential. Determine the correlation coefficient for these data. Assume the data are ordinal in level of measurement.

Sales	Miles per Month
$150,000	1,500
210,000	2,100
285,000	3,200
301,000	2,400
335,000	2,200
390,000	2,500
400,000	3,300
425,000	3,100
440,000	3,600

17.59 Workers in three different but comparable companies were asked to rate the use of quality-control techniques in their firms on a 50-point scale. A score of 50 represents nearly perfect implementation of quality control techniques and 0 represents no implementation. Workers are divided into three independent groups. One group worked in a company that had required all its workers to attend a 3-day seminar on quality control 1 year ago. A second group worked in a company in which each worker was part of a quality circle group that had been meeting at least once a month for a year. The third group of workers was employed by a company in which management had been actively involved in the quality-control process for more than a year. Use $\alpha = .10$ to determine whether there is a significant difference between the three groups, as measured by the ratings.

Attended 3-Day Seminar	Quality Circles	Management Involved
9	27	16
11	38	21
17	25	18
10	40	28
22	31	29
15	19	20
6	35	31

17.60 The scores given are husband-wife scores on a marketing measure. Use the Wilcoxon matched-pairs signed rank test to determine whether the wives' scores are significantly higher on the marketing measure than the husbands'. Assume that $\alpha = .01$.

Husbands	Wives
27	35
22	29
28	30
19	20
28	27
29	31
18	22
21	19
25	29
18	28
20	21
24	22
23	33
25	38
22	34
16	31
23	36
30	31

INTERPRETING THE OUTPUT

17.61 Study the following Minitab output. What statistical test was run? What type of design was it? What was the result of the test?

```
Friedman Test: Observations Versus
Treatment Blocked by Block

S = 11.31  DF = 3   P = 0.010
                              Sum
                      Est      of
Treatment   N   Median   Ranks
1          10   20.125    17.0
2          10   25.875    33.0
3          10   24.500    30.5
4          10   22.500    19.5

Grand median = 23.250
```

17.62 Examine the following Minitab output. Discuss the statistical test, its intent, and its outcome.

```
Runs Test
Runs above and below K = 1.4200

The observed number of runs = 28
The expected number of runs = 25.3600
21 Observations above K, 29 below
P-value = 0.439
```

17.63 Study the following Minitab output. What statistical test was being computed by Minitab? What are the results of this analysis?

```
Mann-Whitney Test and CI

              N   Median
1st Group    16   37.000
2nd Group    16   46.500

Point estimate for ETA1-ETA2 is -8.000
95.7 Percent CI for ETA1-ETA2 is
(-13.999, -2.997)
W = 191.5
Test of ETA1 = ETA2  vs  ETA1 not =
ETA2 is significant at 0.0067
```

17.64 Study the following Minitab output. What type of statistical test was done? What were the hypotheses, and what was the outcome? Discuss.

```
Kruskal-Wallis Test on Observations

Machine     N   Median   Ave Rank       Z
1           5    35.00       14.8    0.82
2           6    25.50        4.2   -3.33
3           7    35.00       15.0    1.11
4           6    35.00       16.0    1.40
Overall    24                12.5
H = 11.21 DF = 3   P = 0.011
```

ANALYZING THE DATABASES

see www.wiley.com/college/black

1. Compute a Spearman's rank correlation between New Capital Expenditures and End-of-Year Inventories in the Manufacture database. Is the amount spent annually on New Capital Expenditures related to the End-of-Year Inventories? Are these two variables highly correlated? Explain.

2. Use a Kruskal-Wallis test to determine whether there is a significant difference between the four levels of Value of Industry Shipments on Number of Employees for the Manufacture database. Discuss the results.

3. Use a Mann-Whitney *U* test to determine whether there is a significant difference between hospitals that are general hospitals and those that are psychiatric (Service variable) on Personnel for the Hospital database. Discuss the results.

4. Use the Kruskal-Wallis test to determine if there is a significant difference in Annual Food Spending by Region in the Consumer Food database.

CASE

SCHWINN

In 1895, Ignaz Schwinn and his partner, Adolph Arnold, incorporated the Arnold, Schwinn & Company in Chicago to produce bicycles. In the early years with bicycle products such as the "Roadster," a single-speed bike that weighed 19 pounds, Schwinn products appealed to people of all ages as a means of transportation. By 1900, bicycles could go as fast as 60 miles per hour. Because of the advent of the auto in 1909, the use of bicycles as a means of transportation in the United States waned. In that same year, Schwinn developed manufacturing advances that allowed bicycles to be made more cheaply and sturdily. These advances opened a new market to the company as they manufactured and sold bicycles for children for the first time. Meanwhile, Ignaz Schwinn bought out Arnold to become the sole owner of the company. Over the next 20 years, Schwinn bought out two motorcycle companies and developed mudguards as its major technological achievement. In the 1930s, Schwinn developed a series of quality, appearance, and technological breakthroughs including the balloon tire, which some say was the biggest innovation in mountain bike technology; the forewheel brake; the cantilever frame; and the spring fork. In 1946, built-in kickstands were added to their bikes. In the 1950s, Schwinn began an authorized dealer network and expanded its parts and accessory programs.

In the 1960s, Schwinn expanded into the fitness arena with in-home workout machines. In 1967, the company became the Schwinn Bicycle Company. The company introduced the airdyne stationary bike in the late 1970s. In 1993, the company filed for bankruptcy; and in 1994, it was moved from Chicago to Boulder, Colorado, to be nearer the mountain bike scene. In the next several years, Schwinn's mountain bike products won accolades and awards. In 2001, Pacific Cycle, the United States' largest importer of quality bicycles, purchased Schwinn and united Schwinn bicycle lines with Pacific Cycle's other brands. Under new management in 2002, Schwinn bicycles began being featured, along with Pacific Cycle's other bikes, at mass retail outlets in the United States. In 2004, Dorel Industries, Inc., a global consumer products company located in Madison, Wisconsin, purchased Pacific Cycle and made it a division of Dorel. Schwinn bicycles, now a part of the Dorel empire, are still made with quality for dependability and performance, and they continue to lead the industry in innovation.

Discussion

1. What is the age of the target market for Schwinn bikes? One theory is that in locales where mountain bikes are more popular, the mean age of their customers is older than in locales where relatively little mountain biking is done. In an attempt to test this theory, a random sample of Colorado Springs customers is taken along with a random sample of customers in St. Louis. The ages for these customers are given here. The customer is defined as "the person for whom the bike is primarily purchased." The shape of the population distribution of bicycle customer ages is unknown. Analyze the data and discuss the implications for Schwinn manufacturing and sales.

Colorado Springs	St. Louis
29	11
38	14
31	15
17	12
36	14
28	25
44	14
9	11
32	8
23	
35	

2. Suppose for a particular model of bike, the specified weight of a handle bar is 200 grams and Schwinn uses three different suppliers of handle bars. Suppose Schwinn conducts a quality-control study in which handle bars are randomly selected from each supplier and weighed. The results (in grams) are shown next. It is uncertain whether handle bar weight is normally distributed in the population. Analyze the data and discuss what the business implications are to Schwinn.

Supplier 1	Supplier 2	Supplier 3
200.76	197.38	192.63
202.63	207.24	199.68
198.03	201.56	203.07
201.24	194.53	195.18
202.88	197.21	189.11
194.62	198.94	
203.58		
205.41		

3. Quality technicians at Schwinn's manufacturing plant examine their finished products for paint flaws. Paint

inspections are done on a production run of 75 bicycles. The inspection data are coded and the data analyzed using Minitab. If a bicycle's paint job contained no flaws, a 0 is recorded; and if it contained at least one flaw, the code used is a 1. Inspectors want to determine whether the flawed bikes occur in a random fashion or in a nonrandom pattern. Study the Minitab output. Determine whether the flaws occur randomly. Report on the proportion of flawed bikes and discuss the implications of these results to Schwinn's production management.

```
Runs Test: Paint Flaw
Runs above and below K = 0.2533

The observed number of runs = 29
The expected number of runs = 29.3733
19 observations above K, 56 below

P-value = 0.908
```

Source: Adapted from Schwinn, available at http://www.schwinnbike.com/heritage.

USING THE COMPUTER

MINITAB

- Five of the nonparametric statistics presented in this chapter can be accessed using Minitab. For each nonparametric technique: select **Stat** from the menu bar. From the **Stat** pulldown menu, select **Nonparametrics**. From **Nonparametrics** pulldown menu, select the appropriate nonparametric technique from **Runs Test, Mann-Whitney**, **Kruskal-Wallis**, **Friedman**, and **1-Sample Wilcoxon**.
- To begin Runs test select **Runs Test** from the **Nonparametrics** pulldown menu. Supply the location of the column with the data in the **Variables** space: Check either **Above and below the mean** or **Above and below**. Minitab will default to **Above and below the mean** and will use the mean of the numbers to determine when the runs stop. Select **Above and below** if you want to supply your own value.
- To begin a Mann-Whitney U test, select **Mann-Whitney** from the **Nonparametrics** pulldown menu. Place the column location of the data from the first sample in **First Sample.** Place the column location of the data from the second sample in **Second Sample.** Insert the level of confidence in **Confidence level.** In the slot provided beside **Alternative**, select the form of the alternative hypothesis. Choose from **not equal, less than**, and **greater than**.
- To begin a Kruskal-Wallis test, all observations from all treatment levels must be located in one column. Place the location of the column containing these observations in the space provided beside **Response**. Enter the treatment levels to match the observations in a second column. Place the location of the column containing these treatment levels in the space provided beside **Factor**.
- To begin a Friedman test, all observations must be located in one column. Place the location of the column containing these observations in the space provided beside **Response**. Enter the treatment levels to match the observations in a second column. Place the location of the column containing these treatment levels in the space provided beside **Treatment**. Enter the block levels to match the observations in a third column. Place the location of the column containing these block levels in the space provided beside **Blocks**. You have the option of storing residuals and storing fits by clicking on **Store residuals** and **Store fits**. You are not allowed to store the fits without storing the residuals.
- There is no Wilcoxin matched-pairs signed rank test in Minitab. You must manipulate Minitab to perform this test. Either enter the data from the two related samples in two columns and use the calculator under **Calc** on the main menu bar to subtract the two columns, thereby creating a third column of differences; or enter the differences in a column to begin with. Select **1-sample Wilcoxon**. In space provided next to **Variables**, enter the location of the column containing the differences. Place the level of confidence in the box beside **Confidence interval**. Minitab will default to 95%. Minitab will default a hypothesis test of a 0.0 median. To test any other value, check **Test median** and enter the new test value for the median. In the slot provided beside **Alternative**, select the form of the alternative hypothesis. Choose from **not equal, less than**, and **greater than**.