# CHAPTER 19

# Nonparametric Methods

## WEST SHELL REALTORS*
*CINCINNATI, OHIO*

West Shell Realtors was founded in 1958 with one office and a sales staff of three people. In 1964, the company began a long-term expansion program, with new offices added almost yearly. Over the years, West Shell grew to become one of the largest realtors in Greater Cincinnati, with offices in southwest Ohio, southeast Indiana, and northern Kentucky.

Statistical analysis helps real estate firms such as West Shell monitor sales performance. Monthly reports are generated for each of West Shell's offices as well as for the total company. Statistical summaries of total sales dollars, number of units sold, and median selling price per unit are essential in keeping both office managers and the company's top management informed of progress and trouble spots in the organization.

In addition to monthly summaries of ongoing operations, the company uses statistical considerations to guide corporate plans and strategies. West Shell has implemented a strategy of planned expansion. Each time an expansion plan calls for the establishment of a new sales office, the company must address the question of office location. Selling prices of homes, turnover rates, and forecast sales volumes are the types of data used in evaluating and comparing alternative locations.

In one instance, West Shell identified two suburbs, Clifton and Roselawn, as prime candidates for a new office. A variety of factors were considered in comparing the two areas, including selling prices of homes. West Shell



West Shell uses statistical analysis of home sales to remain competitive. © Courtesy of Coldwell Banker West Shell.

employed nonparametric statistical methods to help identify any differences in sales patterns for the two areas.

Samples of 25 sales in the Clifton area and 18 sales in the Roselawn area were taken, and the Mann-Whitney-Wilcoxon rank-sum test was chosen as an appropriate statistical test of the difference in the pattern of selling prices. At the .05 level of significance, the Mann-Whitney-Wilcoxon test did not allow rejection of the null hypothesis that the two populations of selling prices were identical. Thus, West Shell was able to focus on criteria other than selling prices of homes in the site selection process.

In this chapter we will show how nonparametric statistical tests such as the Mann-Whitney-Wilcoxon test are applied. We will also discuss the proper interpretation of such tests.

*The authors are indebted to Rodney Fightmaster of West Shell Realtors for providing this Statistics in Practice.

The statistical methods for inference presented previously in the text are generally known as **parametric methods**. These methods begin with an assumption about the probability distribution of the population which is often that the population has a normal distribution. Based upon this assumption, statisticians are able to derive the sampling distribution that can be used to make inferences about one or more parameters of the population, such as the population mean $\mu$ or the population standard deviation $\sigma$. For example, in Chapter 9 we presented a method for making an inference about a population mean that was based on an assumption that the population had a normal probability distribution with unknown parameters $\mu$ and $\sigma$. Using the sample standard deviation $s$ to estimate the population

standard deviation $\sigma$, the test statistic for making an inference about the population mean was shown to have a $t$ distribution. As a result, the $t$ distribution was used to compute confidence intervals and conduct hypothesis tests about the mean of a normally distributed population.

In this chapter we present **nonparametric methods** which can be used to make inferences about a population without requiring an assumption about the specific form of the population's probability distribution. For this reason, these nonparametric methods are also called **distribution-free methods**.

Most of the statistical methods referred to as parametric methods require quantitative data, while nonparametric methods allow inferences based on either categorical or quantitative data. However, the computations used in the nonparametric methods are generally done with categorical data. Whenever the data are quantitative, we will transform the data into categorical data in order to conduct the nonparametric test. In the first section of the chapter, we show how the binomial distribution uses two categories of data to make an inference about a population median. In the next three sections, we show how rank-ordered data are used in nonparametric tests about two or more populations. In the final section, we use rank-ordered data to compute the rank correlation for two variables.

## 19.1   Sign Test

The **sign test** is a versatile nonparametric method for hypothesis testing that uses the binomial distribution with $p = .50$ as the sampling distribution. It does not require an assumption about the distribution of the population. In this section we present two applications of the sign test: one involving a hypothesis test about a population median and one involving a matched-sample test about the difference between two populations.

### Hypothesis Test About a Population Median

In Chapter 9, we described how to conduct hypothesis tests about a population mean. In this section, we show how the sign test can be used to conduct a hypothesis test about a population median. If we consider a population where no data value is exactly equal to the median, the median is the measure of central tendency that divides the population so that 50% of the values are greater than the median and 50% of the values are less than the median. Whenever a population distribution is skewed, the median is often preferred over the mean as the best measure of central location for the population. The sign test provides a nonparametric procedure for testing a hypothesis about the value of a population median.

In order to demonstrate the sign test, we consider the weekly sales of Cape May Potato Chips by the Lawler Grocery Store chain. Lawler's management made the decision to carry the new potato chip product based on the manufacturer's estimate that the median sales should be $450 per week on a per store basis. After carrying the product for three-months, Lawler's management requested the following hypothesis test about the population median weekly sales.

$$H_0: \text{Median} = 450$$
$$H_a: \text{Median} \neq 450$$

Data showing one-week sales at 10 randomly selected Lawler's stores are provided in Table 19.1.

**TABLE 19.1**    ONE-WEEK SALES OF CAPE MAY POTATO CHIPS AT 10 LAWLER
GROCERY STORES

| Store Number | Weekly Sales ($) | Store Number | Weekly Sales ($) |
|---|---|---|---|
| 56 | 485 | 63 | 474 |
| 19 | 562 | 39 | 662 |
| 36 | 415 | 84 | 380 |
| 128 | 860 | 102 | 515 |
| 12 | 426 | 44 | 721 |

*Observations equal to the hypothesized value are discarded and the analysis proceeds with the observations having either a plus sign or a minus sign.*

In conducting the sign test, we compare each sample observation to the hypothesized value of the population median. If the observation is greater than the hypothesized value, we record a plus sign "$+$." If the observation is less than the hypothesized value, we record a minus sign "$-$." If an observation is exactly equal to the hypothesized value, the observation is eliminated from the sample and the analysis proceeds with the smaller sample size, using only the observations where a plus sign or a minus sign has been recorded. It is the conversion of the sample data to either a plus sign or a minus sign that gives the nonparametric method its name: the sign test.

Consider the sample data in Table 19.1. The first observation, 485, is greater than the hypothesized median 450; a plus sign is recorded. The second observation, 562, is greater than the hypothesized median 450; a plus sign is recorded. Continuing with the 10 observations in the sample provides the plus and minus signs as shown in Table 19.2. Note that there are 7 plus signs and 3 minus signs.

The assigning of the plus signs and minus signs has made the situation a binomial distribution application. The sample size $n = 10$ is the number of trials. There are two outcomes possible per trial, a plus sign or a minus sign, and the trials are independent. Let $p$ denote the probability of a plus sign. If the population median is 450, $p$ would equal .50 as there should be 50% plus signs and 50% minus signs in the population. Thus, in terms of the binomial probability $p$, the sign test hypotheses about the population median

$$H_0\text{: Median} = 450$$
$$H_a\text{: Median} \neq 450$$

are converted to the following hypotheses about the binomial probability $p$.

$$H_0\text{: } p = .50$$
$$H_a\text{: } p \neq .50$$

**TABLE 19.2**    LAWLER SAMPLE DATA FOR THE SIGN TEST ABOUT THE POPULATION
MEDIAN WEEKLY SALES

| Store Number | Weekly Sales ($) | Sign | Store Number | Weekly Sales ($) | Sign |
|---|---|---|---|---|---|
| 56 | 485 | + | 63 | 474 | + |
| 19 | 562 | + | 39 | 662 | + |
| 36 | 415 | − | 84 | 380 | − |
| 128 | 860 | + | 102 | 515 | + |
| 12 | 426 | − | 44 | 721 | + |

**TABLE 19.3**

BINOMIAL
PROBABILITIES
WITH $n = 10$ AND
$p = .50$

| Number of Plus Signs | Probability |
|---|---|
| 0 | .0010 |
| 1 | .0098 |
| 2 | .0439 |
| 3 | .1172 |
| 4 | .2051 |
| 5 | .2461 |
| 6 | .2051 |
| 7 | .1172 |
| 8 | .0439 |
| 9 | .0098 |
| 10 | .0010 |

*Binomial probabilities are provided in Table 5 of Appendix B when the sample size is less than or equal to 20. Excel or Minitab can be used to provide binomial probabilities for any sample size.*

If $H_0$ cannot be rejected, we cannot conclude that $p$ is different from .50 and thus we cannot conclude that the population median is different from 450. However, if $H_0$ is rejected, we can conclude that $p$ is not equal to .50 and thus the population median is not equal to 450.

With $n = 10$ stores or trials and $p = .50$, we used Table 5 in Appendix B to obtain the binomial probabilities for the number of plus signs under the assumption $H_0$ is true. These probabilities are shown in Table 19.3. Figure 19.1 shows a graphical representation of this binomial distribution.

Let us proceed to show how the binomial distribution can be used to test the hypothesis about the population median. We will use a .10 level of significance for the test. Since the observed number of plus signs for the sample data, 7, is in the upper tail of the binomial distribution, we begin by computing the probability of obtaining 7 or more plus signs. This probability is the probability of 7, 8, 9, or 10 plus signs. Adding these probabilities shown in Table 19.3, we have $.1172 + .0439 + .0098 + .0010 = .1719$. Since we are using a two-tailed hypothesis test, this upper tail probability is doubled to obtain the $p$-value $= 2(.1719) = .3438$. With $p$-value $> \alpha$, we cannot reject $H_0$. In terms of the binomial probability $p$, we cannot reject $H_0$: $p = .50$, and thus we cannot reject the hypothesis that the population median is $450.

In this example, the hypothesis test about the population median was formulated as a two-tailed test. However, one-tailed sign tests about a population median are also possible. For example, we could have formulated the hypotheses as an upper tail test so that the null and alternative hypotheses would be written as follows:

$$H_0: \text{Median} \leq 450$$
$$H_a: \text{Median} > 450$$

The corresponding $p$-value is equal to the binomial probability that the number of plus signs is greater than or equal to 7 found in the sample. This one-tailed $p$-value would have been $.1172 + .0439 + .0098 + .0010 = .1719$. If the example were converted to a lower tail test, the $p$-value would have been the probability of obtaining 7 or fewer plus signs.

**FIGURE 19.1**    BINOMIAL SAMPLING DISTRIBUTION FOR THE NUMBER OF PLUS SIGNS WHEN $n = 10$ AND $p = .50$
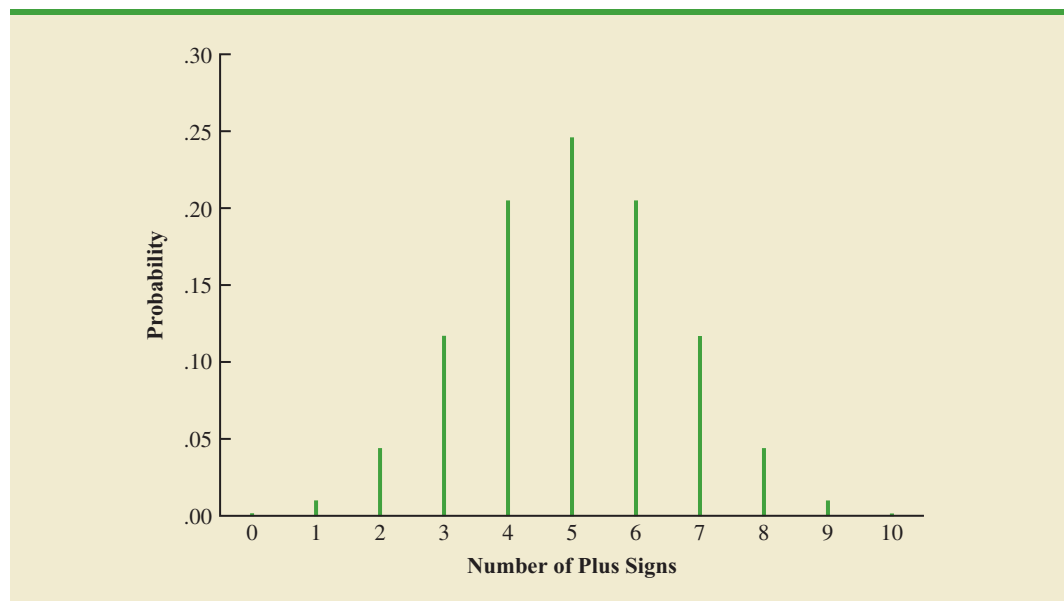
The application we have just described makes use of the binomial distribution with $p = .50$. The binomial probabilities provided in Table 5 of Appendix B can be used to compute the $p$-value when the sample size is 20 or less. With larger sample sizes, we rely on the normal distribution approximation of the binomial distribution to compute the $p$-value; this makes the computations quicker and easier. A large sample application of the sign test is illustrated in the following example.

One year ago the median price of a new home was $236,000. However, a current downturn in the economy has real estate firms using sample data on recent home sales to determine if the population median price of a new home is less today than it was a year ago. The hypothesis test about the population median price of a new home is as follows:

$$H_0\text{: Median} \geq 236{,}000$$
$$H_a\text{: Median} < 236{,}000$$

**WEB file**

**HomeSales**

We will use a .05 level of significance to conduct this test.

A random sample of 61 recent new home sales found 22 homes sold for more than $236,000, 38 homes sold for less than $236,000, and one home sold for $236,000. After deleting the home that sold for the hypothesized median price of $236,000, the sign test continues with 22 plus signs, 38 minus signs, and a sample of 60 homes.

The null hypothesis that the population median is greater than or equal to $236,000 is expressed by the binomial distribution hypothesis $H_0\text{: } p \geq .50$. If $H_0$ were true as an equality, we would expect $.50(60) = 30$ homes to have a plus sign. The sample result showing 22 plus signs is in the lower tail of the binomial distribution. Thus, the $p$-value is the probability of 22 or fewer plus signs when $p = .50$. While it is possible to compute the exact binomial probabilities for 0, 1, 2, . . . to 22 and sum these probabilities, we will use the normal distribution approximation of the binomial distribution to make this computation easier. For this approximation, the mean and standard deviation of the normal distribution are as follows.

---

NORMAL APPROXIMATION OF THE SAMPLING DISTRIBUTION OF THE
NUMBER OF PLUS SIGNS WHEN
$H_0\text{: } p = .50$

$$\text{Mean: } \mu = .50n \qquad\qquad\textbf{(19.1)}$$
$$\text{Standard deviation: } \sigma = \sqrt{.25n} \qquad\qquad\textbf{(19.2)}$$

Distribution form: Approximately normal for $n > 20$

---

Using equations (19.1) and (19.2) with $n = 60$ homes and $p = .50$, the sampling distribution of the number of plus signs can be approximated by a normal distribution with

$$\mu = .50n = .50(60) = 30$$
$$\sigma = \sqrt{.25n} = \sqrt{.25(60)} = 3.873$$

Let us now use the normal distribution to approximate the binomial probability of 22 or fewer plus signs. Before we proceed, remember that the binomial probability distribution is discrete and the normal probability distribution is continuous. To account for this, the binomial probability of 22 is computed by the normal probability interval 21.5 to 22.5. The .5 added to and subtracted from 22 is called the continuity correction factor. Thus, to compute

**FIGURE 19.2**    NORMAL DISTRIBUTION APPROXIMATION OF THE $p$-VALUE FOR THE
SIGN TEST ABOUT THE MEDIAN PRICE OF NEW HOMES
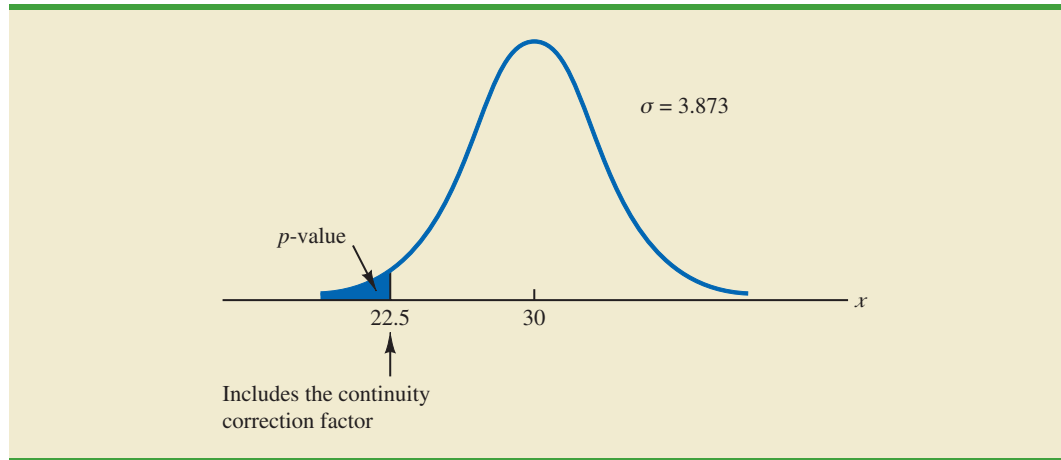


the $p$-value for 22 or fewer plus signs we use the normal distribution with $\mu = 30$ and
$\sigma = 3.873$ to compute the probability that the normal random variable, $x$, has a value less
than or equal to 22.5. A graph of this $p$-value is shown in Figure 19.2.

Using this normal distribution, we compute the $p$-value as follows:

$$p\text{-value} = P(x \le 22.5) = P\left(z \le \frac{22.5 - 30}{3.873}\right) = P(z \le -1.94)$$

Using the table of areas for a normal probability distribution, we see that the cumulative
probability for $z = -1.94$ provides the $p$-value $= .0262$. With $.0262 < .05$, we reject the null
hypothesis and conclude that the median price of a new home is less than the \$236,000
median price a year ago.

## NOTES AND COMMENTS

1. The examples used to illustrate a hypothesis test
about a population median involved weekly
sales data and home price data. The probability
distributions for these types of variables are
usually not symmetrical and are most often
skewed to the right. In such cases, the popula-
tion median rather than the population mean
becomes the preferred measure of central loca-
tion. In general, when the population is not sym-
metrical, the nonparametric sign test for the
population median is often the more appropriate
statistical test.

2. The binomial sampling distribution for the sign
test can be used to compute a confidence inter-
val estimate of the population median. However,
the computations are rather complex and would
rarely be done by hand. Statistical packages such
as Minitab can be used to obtain a confidence in-
terval for a population median. The Minitab pro-
cedure to do this is described in Appendix 19.1.
Using the price of homes example in this section,
Minitab provides the 95% confidence interval
for the median price of a new home as \$183,000
to \$231,000.

## Hypothesis Test with Matched Samples

In Chapter 10, we introduced a matched-sample experimental design where each of $n$ experimental units provided a pair of observations, one from population 1 and one from population 2. Using quantitative data and assuming that the differences between the pairs of matched observations were normally distributed, the $t$ distribution was used to make an inference about the difference between the means of the two populations.

In the following example we will use the nonparametric sign test to analyze matched-sample data. Unlike the $t$ distribution procedure, which required quantitative data and the assumption that the differences were normally distributed, the sign test enables us to analyze categorical as well as quantitative data and requires no assumption about the distribution of the differences. This type of matched-sample design occurs in market research when a sample of $n$ potential customers is asked to compare two brands of a product such as coffee, soft drinks, or detergents. Without obtaining a quantitative measure of each individual's preference for the brands, each individual is asked to state a brand preference. Consider the following example.

Sun Coast Farms produces an orange juice product called Citrus Valley. The primary competition for Citrus Valley comes from the producer of an orange juice known as Tropical Orange. In a consumer preference comparison of the two brands, 14 individuals were given unmarked samples of the two orange juice products. The brand each individual tasted first was selected randomly. If the individual selected Citrus Valley as the more preferred, a plus sign was recorded. If the individual selected Tropical Orange as the more preferred, a minus sign was recorded. If the individual was unable to express a difference in preference for the two products, no sign was recorded. The data for the 14 individuals in the study are shown in Table 19.4.

Deleting the two individuals who could not express a preference for either brand, the data have been converted to a sign test with 2 plus signs and 10 minus signs for the $n = 12$ individuals who could express a preference for one of the two brands. Letting $p$ indicate the proportion of the population of customers who prefer Citrus Valley orange juice, we want to test the hypotheses that there is no difference between the preferences for the two brands as follows:

$$H_0: p = .50$$
$$H_a: p \neq .50$$

If $H_0$ cannot be rejected, we cannot conclude that there is a difference in preference for the two brands. However, if $H_0$ can be rejected, we can conclude that the consumer preferences differ for the two brands. We will use a .05 level of significance for this hypothesis test.

We will conduct the sign test exactly as we did earlier in this section. The sampling distribution for the number of plus signs is a binomial distribution with $p = .50$ and $n = 12$.

**TABLE 19.4**   PREFERENCE DATA FOR THE SUN COAST FARMS TASTE TEST

| Individual | Preference | Sign | Individual | Preference | Sign |
|---|---|---|---|---|---|
| 1 | Tropical Orange | − | 8 | Tropical Orange | − |
| 2 | Tropical Orange | − | 9 | Tropical Orange | − |
| 3 | Citrus Valley | + | 10 | No Preference | |
| 4 | Tropical Orange | − | 11 | Tropical Orange | − |
| 5 | Tropical Orange | − | 12 | Citrus Valley | + |
| 6 | No Preference | | 13 | Tropical Orange | − |
| 7 | Tropical Orange | − | 14 | Tropical Orange | − |

Using Table 5 in Appendix B we obtain the binomial probabilities for the number of plus signs, as shown in Table 19.5. Under the assumption $H_0$ is true, we would expect $.50n = .50(12) = 6$ plus signs. With only two plus signs in the sample, the results are in the lower tail of the binomial distribution. To compute the $p$-value for this two-tailed test, we first compute the probability of 2 or fewer plus signs and then double this value. Using the binomial probabilities of 0, 1, and 2 shown in Table 19.5, the $p$-value is $2(.0002 + .0029 + .0161) = .0384$. With $.0384 < .05$, we reject $H_0$. The taste test provides evidence that consumer preference differs significantly for the two brands of orange juice. We would advise Sun Coast Farms of this result and conclude that the competitor's Tropical Orange product is the more preferred. Sun Coast Farms can then pursue a strategy to address this issue.

Similar to other uses of the sign test, one-tailed tests may be used depending upon the application. Also, as the sample size becomes large, the normal distribution approximation of the binomial distribution will ease the computations as shown earlier in this section. While the Sun Coast Farms sign test for matched samples used categorical preference data, the sign test for matched samples can be used with quantitative data as well. This would be particularly helpful if the paired differences are not normally distributed and are skewed. In this case a positive difference is assigned a plus sign, a negative difference is assigned a negative sign, and a zero difference is removed from the sample. The sign test computations proceed as before.

**TABLE 19.5**

BINOMIAL PROBABILITIES WITH $n = 12$ AND $p = .50$

| Number of Plus Signs | Probability |
|---|---|
| 0 | .0002 |
| 1 | .0029 |
| 2 | .0161 |
| 3 | .0537 |
| 4 | .1208 |
| 5 | .1934 |
| 6 | .2256 |
| 7 | .1934 |
| 8 | .1208 |
| 9 | .0537 |
| 10 | .0161 |
| 11 | .0029 |
| 12 | .0002 |

## Exercises

### Methods

**SELF test**

1. The following hypothesis test is to be conducted.

$$H_0: \text{Median} \geq 150$$
$$H_a: \text{Median} < 150$$

A sample of 30 provided 22 observations greater than 150, 3 observations equal to 150, and 5 observations less than 150. Use $\alpha = .01$. What is your conclusion?

**SELF test**

2. Ten individuals participated in a taste test involving two brands of a product. Sample results show 7 preferred brand A, 2 preferred brand B, and 1 was unable to state a preference. With $\alpha = .05$, test for a significant difference in the preferences for the two brands. What is your conclusion?

### Applications

3. The median number of part-time employees at fast-food restaurants in a particular city was known to be 18 last year. City officials think the use of part-time employees may be increasing. A sample of nine fast-food restaurants showed that seven restaurants were employing more than 18 part-time employees, one restaurant was employing exactly 18 part-time employees, and one restaurant was employing fewer than 18 part-time employees. Can it be concluded that the median number of part-time employees has increased? Test using $\alpha = .05$.

4. Net assets for the 50 largest stock mutual funds show a median of $15 billion (*The Wall Street Journal*, March 2, 2009). A sample of 10 of the 50 largest bond mutual funds follows.

| Bond Fund | Net Assets | Bond Fund | Net Assets |
|---|---|---|---|
| Fidelity Intl Bond | 6.1 | T Rowe Price New Income | 6.9 |
| Franklin CA TF | 11.7 | Vanguard GNMA | 15.0 |
| American Funds | 22.4 | Oppenheimer Intl Bond | 6.6 |
| Vanguard Short Term | 9.6 | Dodge & Cox Income | 14.5 |
| PIMCO: Real Return | 4.9 | iShares: TIPS Bond | 9.6 |

Using the median, can it be concluded that bond mutual funds are smaller and have fewer net assets than stock mutual funds? Use $\alpha = .05$.
a. What are the hypotheses for this test?
b. What is the *p*-value? What is your conclusion?

5. The median annual income of subscribers to *Shutterbug* magazine is $75,000 (Home Theater website, August 18, 2008). A sample of 300 subscribers to *Popular Photography & Imaging* magazine found 165 subscribers with an annual income over $75,000 and 135 with an annual income under $75,000. Can you conclude that the median annual income of *Popular Photography & Imaging* subscribers differs from the median annual income of *Shutterbug* subscribers? Use $\alpha = .05$.

6. The median annual income for families living in the United States is $56,200 (*The New York Times Almanac,* 2008). Annual incomes in thousands of dollars for a sample of 50 families living in Chicago, Illinois, are shown. Use the sample data to see if it can be concluded that the families living in Chicago have a median annual income greater than $56,200. Use $\alpha = .05$. What is your conclusion?

**WEB file**

**ChicagoIncome**

| | | | | |
|---|---|---|---|---|
| 66.3 | 60.2 | 49.9 | 75.4 | 73.7 |
| 65.7 | 61.1 | 123.8 | 57.3 | 48.5 |
| 74.0 | 146.3 | 92.2 | 43.7 | 86.9 |
| 59.7 | 64.2 | 56.2 | 48.9 | 109.6 |
| 39.8 | 60.9 | 79.7 | 42.3 | 52.6 |
| 60.9 | 43.5 | 61.7 | 54.7 | 95.2 |
| 70.4 | 43.8 | 57.8 | 83.5 | 56.5 |
| 51.3 | 42.9 | 87.5 | 43.6 | 67.2 |
| 48.7 | 79.1 | 61.9 | 53.4 | 56.2 |
| 57.0 | 49.6 | 109.5 | 42.1 | 74.6 |

7. Are stock splits beneficial to stockholders? SNL Financial studied stock splits in the banking industry over an 18-month period. For a sample of 20 stock splits, 14 led to an increase in investment value, 4 led to a decrease in investment value, and 2 resulted in no change. Conduct a sign test to determine if it can be concluded that stock splits are beneficial for holders of bank stocks.
a. What are the null and alternative hypotheses?
b. Using $\alpha = .05$, what is your conclusion?

8. A Pew Research Center survey asked adults if their ideal place to live would have a faster pace of life or a slower pace of life (*USA Today*, February 13, 2009). A preliminary sample of 16 respondents showed 4 preferred a faster pace of life, 11 preferred a slower place of life, and 1 said it did not matter.
a. Are these data sufficient to conclude there is a difference between the preferences for a faster pace of life or a slower pace of life? Use $\alpha = .05$. What is your conclusion?
b. Considering the entire sample of 16 respondents, what is the percentage who would like a faster pace of life? What is the percentage who would like a slower pace of life? What recommendation do you have for the study?

9. A poll taken during the recession in 2008 asked 600 adults a series of questions about the state of the economy and their children's future. One question was, "Do you expect your children to have a better life than you have had, a worse life, or a life about the same as yours?" The responses showed 242 better, 310 worse, and 48 about the same. Use the sign test and $\alpha = .05$ to determine whether there is a difference between the number of adults who feel their children will have a better life compared to a worse life. What is your conclusion?

10.  Nielsen Media Research identified *American Idol* and *Dancing with the Stars* as the two top-rated prime-time television shows (*USA Today*, April 14, 2008). In a local television preference survey, 750 individuals were asked to indicate their favorite prime-time television show: Three hundred thirty selected *American Idol*, 270 selected *Dancing with the Stars*, and 150 selected another television show. Use a .05 level of significance to test the hypothesis that there is no difference in the preference for the *American Idol* and *Dancing with the Stars* television shows. What is your conclusion?

11.  Competition in the personal computer market is intense. A sample of 450 purchases showed 202 Brand A computers, 175 Brand B computers, and 73 other computers. Use a .05 level of significance to test the null hypothesis that Brand A and Brand B have the same share of the personal computer market. What is your conclusion?

## 19.2   Wilcoxon Signed–Rank Test

In Chapter 10, we introduced a matched-sample experimental design where each of *n* experimental units provided a pair of observations, one from population 1 and one from population 2. The parametric test for this experiment requires quantitative data and the assumption that the differences between the paired observations are normally distributed. The *t* distribution can then be used to make an inference about the difference between the means of the two populations.

*If the population of differences is skewed, the sign test for matched samples presented in Section 19.1 is recommended.*

The **Wilcoxon signed-rank test** is a nonparametric procedure for analyzing data from a matched-sample experiment. The test uses quantitative data but does not require the assumption that the differences between the paired observations are normally distributed. It only requires the assumption that the differences between the paired observations have a symmetric distribution. This occurs whenever the shapes of the two populations are the same and the focus is on determining if there is a difference between the medians of the two populations. Let us demonstrate the Wilcoxon signed-rank test with the following example.

Consider a manufacturing firm that is attempting to determine whether two production methods differ in terms of task completion time. Using a matched-samples experimental design, 11 randomly selected workers completed the production task two times, once using method A and once using method B. The production method that the worker used first was randomly selected. The completion times for the two methods and the differences between the completion times are shown in Table 19.6. A positive difference indicates that method

**TABLE 19.6**   PRODUCTION TASK COMPLETION TIMES (MINUTES)

| Worker | Method A | Method B | Difference |
|--------|------|------|------------|
| 1 | 10.2 | 9.5 | .7 |
| 2 | 9.6 | 9.8 | −.2 |
| 3 | 9.2 | 8.8 | .4 |
| 4 | 10.6 | 10.1 | .5 |
| 5 | 9.9 | 10.3 | −.4 |
| 6 | 10.2 | 9.3 | .9 |
| 7 | 10.6 | 10.5 | .1 |
| 8 | 10.0 | 10.0 | .0 |
| 9 | 11.2 | 10.6 | .6 |
| 10 | 10.7 | 10.2 | .5 |
| 11 | 10.6 | 9.8 | .8 |

A required more time; a negative difference indicates that method B required more time. Do the data indicate that the two production methods differ significantly in terms of completion times? If we assume that the differences have a symmetric distribution but not necessarily a normal distribution, the Wilcoxon signed-rank test applies.

*The examples in this section take the point of view that the two populations have the same shape and if they do differ, it is only in location. This enables the hypotheses for the Wilcoxon signed-rank test to be stated in terms of the population medians.*

In particular, we will use the Wilcoxon signed-rank test for the difference between the median completion times for the two production methods. The hypotheses are as follows:

$$H_0: \text{Median for method A} - \text{Median for method B} = 0$$
$$H_a: \text{Median for method A} - \text{Median for method B} \neq 0$$

If $H_0$ cannot be rejected, we will not be able to conclude that the median completion times are different. However, if $H_0$ is rejected, we will conclude that the median completion times are different. We will use a .05 level of significance for the test.

*Differences of 0 are discarded and the analysis continues with the smaller sample size involving the nonzero differences.*

The first step in the Wilcoxon signed-rank test is to discard the difference of zero for worker 8 and then compute the absolute value of the differences for the remaining 10 workers as shown in column 3 of Table 19.7. Next we rank these absolute differences from lowest to highest as shown in column 4. The smallest absolute difference of .1 for worker 7 is assigned the rank of 1. The second smallest absolute difference of .2 for worker 2 is assigned the rank of 2. This ranking of absolute differences continues with the largest absolute difference of .9 for worker 6 being assigned the rank of 10. The tied absolute differences of .4 for workers 3 and 5 are assigned the average rank of 3.5. Similarly, the tied absolute differences of .5 for workers 4 and 10 are assigned the average rank of 5.5.

*Ties among absolute differences are assigned the average of their ranks.*

Once the ranks of the absolute differences have been determined, each rank is given the *sign* of the original difference for the worker. The negative signed ranks are placed in column 5 and the positive signed ranks are placed in column 6 (see Table 19.7). For example, the difference for worker 1 was a positive .7 (see column 2) and the rank of the absolute difference was 8 (see column 4). Thus, the rank for worker 1 is shown as a positive signed rank in column 6. The difference for worker 2 was a negative .2 and the rank of the absolute difference was 2. Thus, the rank for worker 2 is shown as a negative signed rank of $-2$ in column 5. Continuing this process generates the negative and positive signed ranks as shown in Table 19.7.

**TABLE 19.7** RANKING THE ABSOLUTE DIFFERENCES AND THE SIGNED RANKS FOR THE PRODUCTION TASK COMPLETION TIMES

| Worker | Difference | Absolute Difference | Rank | Signed Ranks Negative | Signed Ranks Positive |
|--------|-----------|---------------------|------|------------------|------------------|
| 1 | .7 | .7 | 8 | | 8 |
| 2 | −.2 | .2 | 2 | −2 | |
| 3 | .4 | .4 | 3.5 | | 3.5 |
| 4 | .5 | .5 | 5.5 | | 5.5 |
| 5 | −.4 | .4 | 3.5 | −3.5 | |
| 6 | .9 | .9 | 10 | | 10 |
| 7 | .1 | .1 | 1 | | 1 |
| 8 | .0 | | | | |
| 9 | .6 | .6 | 7 | | 7 |
| 10 | .5 | .5 | 5.5 | | 5.5 |
| 11 | .8 | .8 | 9 | | 9 |

Sum of Positive Signed Ranks $T^+ = 49.5$

Let $T^+$ denote the sum of the positive signed ranks, which is $T^+ = 49.5$. To conduct the Wilcoxon signed-rank test, we will use $T^+$ as the test statistic. If the medians of the two populations are equal and the number of matched pairs is 10 or more, the sampling distribution of $T^+$ can be approximated by a normal distribution as follows.

SAMPLING DISTRIBUTION OF $T^+$ FOR THE WILCOXON SIGNED-RANK TEST

$$\text{Mean: } \mu_{T^+} = \frac{n(n + 1)}{4} \qquad \textbf{(19.3)}$$

$$\text{Standard deviation: } \sigma_{T^+} = \sqrt{\frac{n(n + 1)(2n + 1)}{24}} \qquad \textbf{(19.4)}$$

Distribution Form: Approximately normal for $n \geq 10$

After discarding the observation of a zero difference for worker 8, the analysis continues with the $n = 10$ matched pairs. Using equations (19.3) and (19.4), we have

$$\mu_{T^+} = \frac{n(n + 1)}{4} = \frac{10(10 + 1)}{4} = 27.5$$

$$\sigma_{T^+} = \sqrt{\frac{n(n + 1)(2n + 1)}{24}} = \sqrt{\frac{10(10 + 1)(20 + 1)}{24}} = \sqrt{\frac{2310}{24}} = 9.8107$$

Figure 19.3 shows the sampling distribution of the $T^+$ test statistic.

Let us compute the two-tailed $p$-value for the hypothesis that the median completion times for the two production methods are equal. Since the test statistic $T^+ = 49.5$ is in the upper tail of the sampling distribution, we begin by computing the upper tail probability $P(T^+ \geq 49.5)$. Since the sum of the positive ranks $T^+$ is discrete and the normal distribution is continuous, we will obtain the best approximation by including the continuity

**FIGURE 19.3**   SAMPLING DISTRIBUTION OF $T^+$ FOR THE PRODUCTION TASK
COMPLETION TIME EXAMPLE
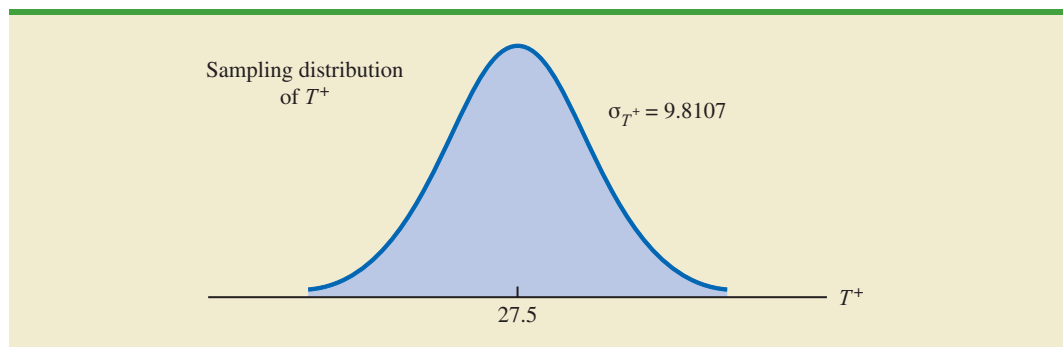
correction factor. Thus, the discrete probability of $T^+ = 49.5$ is approximated by the normal probability interval, 49 to 50, and the probability that $T^+ \geq 49.5$ is approximated by:

$$P(T^+ \geq 49.5) = P\left(z \geq \frac{49 - 27.5}{9.8107}\right) = P(z \geq 2.19)$$

Using the standard normal distribution table and $z = 2.19$, we see that the two-tailed $p$-value $= 2(1-.9857) = .0286$. With the $p$-value $\leq .05$, we reject $H_0$ and conclude that the median completion times for the two production methods are not equal. With $T^+$ being in the upper tail of the sampling distribution, we see that method A led to the longer completion times. We would expect management to conclude that method B is the faster or better production method.

One-tailed Wilcoxon signed-rank tests are possible. For example, if initially we had been looking for statistical evidence to conclude method A had the larger median completion time and method B has the smaller median completion time, we would have formulated an upper tail hypothesis test as follows:

$$H_0: \text{Median for method A} - \text{Median for method B} \leq 0$$
$$H_a: \text{Median for method A} - \text{Median for method B} > 0$$

Rejecting $H_0$ would provide the conclusion that method A has the greater median completion time and method B has the smaller median completion time. Lower tail hypothesis tests are also possible.

*The Wilcoxon signed-rank test can be used to test the hypothesis about the median of a symmetric population. If the population is skewed, the sign test presented in Section 19.1 is preferred.*

As a final note, in Section 19.1 we showed how the sign test could be used for both a hypothesis test about a population median and a hypothesis test with matched samples. In this section, we have demonstrated the use of the Wilcoxon signed-rank test for a hypothesis test with matched samples. However, the Wilcoxon signed-rank test can also be used for a nonparametric test about a population median. This test makes no assumption about the population distribution other than that it is symmetric. If this assumption is appropriate, the Wilcoxon signed-rank test is the preferred nonparametric test for a population median. However, if the population is skewed, the sign test presented in Section 19.1 is preferred. With the Wilcoxon signed-rank test, the differences between the observations and the hypothesized value of the population median are used instead of the differences between the matched-pair observations. Otherwise the calculations are exactly as shown in this section. Exercise 17 will ask you to use the Wilcoxon signed-rank test to conduct a hypothesis test about the median of a symmetric population.

## NOTES AND COMMENTS

1. The Wilcoxon signed-rank test for a population median is based on the assumption that the population is symmetric. With this assumption, the population median is equal to the population mean. Thus, the Wilcoxon signed-rank test can also be used as a test about the mean of a symmetric population.

2. The Wilcoxon signed-rank procedure can also be used to compute a confidence interval for the median of a symmetric population. However, the computations are rather complex and would rarely be done by hand. Statistical packages such as Minitab can be used to obtain this confidence interval.

## Exercises

## Applications

In the following exercises involving paired differences, consider that it is reasonable to assume the populations being compared have approximately the same shape and that the distribution of paired differences is approximately symmetric.

**SELF** test

**WEB** file

Additive

12. Two fuel additives are tested to determine their effect on miles per gallon for passenger cars. Test results for 12 cars follow; each car was tested with both fuel additives. Use $\alpha = .05$ and the Wilcoxon signed-rank test to see whether there is a significant difference between the median miles per gallon for the additives.

| | Additive | | | Additive | |
|---|---|---|---|---|---|
| **Car** | **1** | **2** | **Car** | **1** | **2** |
| 1 | 20.12 | 18.05 | 7 | 16.16 | 17.20 |
| 2 | 23.56 | 21.77 | 8 | 18.55 | 14.98 |
| 3 | 22.03 | 22.57 | 9 | 21.87 | 20.03 |
| 4 | 19.15 | 17.06 | 10 | 24.23 | 21.15 |
| 5 | 21.23 | 21.22 | 11 | 23.21 | 22.78 |
| 6 | 24.77 | 23.80 | 12 | 25.02 | 23.70 |

**SELF** test

13. A sample of 10 men was used in a study to test the effects of a relaxant on the time required to fall asleep. Data for 10 subjects showing the number of minutes required to fall asleep with and without the relaxant follow. Use a .05 level of significance to determine whether the relaxant reduces the median time required to fall asleep. What is your conclusion?

**WEB** file

Relaxant

| | Relaxant | | | Relaxant | |
|---|---|---|---|---|---|
| **Subject** | **No** | **Yes** | **Subject** | **No** | **Yes** |
| 1 | 15 | 10 | 6 | 7 | 5 |
| 2 | 12 | 10 | 7 | 8 | 10 |
| 3 | 22 | 12 | 8 | 10 | 7 |
| 4 | 8 | 11 | 9 | 14 | 11 |
| 5 | 10 | 9 | 10 | 9 | 6 |

14. Percents of on-time arrivals for flights in 2006 and 2007 were collected for 11 randomly selected airports. Data for these airports follow (Research and Innovative Technology Administration website, August 29, 2008). Use $\alpha = .05$ to test the hypothesis that there is no difference between the median percent of on-time arrivals for the two years. What is your conclusion?

|  | | **Percent On Time** |
| **Airport** | **2006** | **2007** |
| Boston Logan | 71.78 | 69.69 |
| Chicago O'Hare | 68.23 | 65.88 |
| Chicago Midway | 77.98 | 78.40 |
| Denver | 78.71 | 75.78 |
| Fort Lauderdale | 77.59 | 73.45 |
| Houston | 77.67 | 78.68 |
| Los Angeles | 76.67 | 76.38 |
| Miami | 76.29 | 70.98 |
| New York (JFK) | 69.39 | 62.84 |
| Orlando | 79.91 | 76.49 |
| Washington (Dulles) | 75.55 | 72.42 |

**WEB** file

**OnTime**

15. A test was conducted for two overnight mail delivery services. Two samples of identical deliveries were set up so that both delivery services were notified of the need for a delivery at the same time. The hours required to make each delivery follow. Do the data shown suggest a difference in the median delivery times for the two services? Use a .05 level of significance for the test.

|  | | **Service** |
| **Delivery** | **1** | **2** |
| 1 | 24.5 | 28.0 |
| 2 | 26.0 | 25.5 |
| 3 | 28.0 | 32.0 |
| 4 | 21.0 | 20.0 |
| 5 | 18.0 | 19.5 |
| 6 | 36.0 | 28.0 |
| 7 | 25.0 | 29.0 |
| 8 | 21.0 | 22.0 |
| 9 | 24.0 | 23.5 |
| 10 | 26.0 | 29.5 |
| 11 | 31.0 | 30.0 |

**WEB** file

**Overnight**

16. The PGA Players Championship was held at the Sedgefield Country Club in Greensboro, North Carolina, August 11–17, 2008. Shown here are first-round and second-round scores for a random sample of 11 golfers. Use $\alpha = .05$ to determine whether the first- and second-round median scores for golfers in the Players Championship differed significantly. What is your conclusion?

| **Golfer** | **1st Round** | **2nd Round** |
| Marvin Laird | 63 | 74 |
| Jimmy Walker | 70 | 73 |
| Kevin Chappell | 72 | 70 |

**WEB** file

**GolfScores**

(*Continued*)

| Golfer | 1st Round | 2nd Round |
|--------|-----------|-----------|
| Kevin Duke | 65 | 71 |
| Andrew Buckle | 70 | 74 |
| Paul Claxton | 69 | 73 |
| Larry Mize | 72 | 71 |
| Chris Riley | 68 | 70 |
| Bubba Watson | 70 | 68 |
| Carlos Franco | 71 | 71 |
| Richard Johnson | 72 | 69 |

17. The Scholastic Aptitude Test (SAT) consists of three parts: critical reading, mathematics, and writing. Each part of the test is scored on a 200- to 800-point scale with a median of approximately 500 (*The World Almanac,* 2009). Scores for each part of the test can be assumed to be symmetric. Use the following data to test the hypothesis that the population median score for the students taking the writing portion of the SAT is 500. Using $\alpha = .05$, what is your conclusion?

**WEB** file

**WritingScore**

| 635 | 701 | 439 | 447 | 464 |
|-----|-----|-----|-----|-----|
| 502 | 405 | 453 | 471 | 476 |
| 447 | 590 | 337 | 387 | 514 |

## 19.3   Mann–Whitney–Wilcoxon Test

In Chapter 10, we introduced a procedure for conducting a hypothesis test about the difference between the means of two populations using two independent samples, one from population 1 and one from population 2. This parametric test required quantitative data and the assumption that both populations had a normal distribution. In the case where the population standard deviations $\sigma_1$ and $\sigma_2$ were unknown, the sample standard deviations $s_1$ and $s_2$ provided estimates of $\sigma_1$ and $\sigma_2$ and the $t$ distribution was used to make an inference about the difference between the means of the two populations.

   In this section we present a nonparametric test for the difference between two populations based on two independent samples. Advantages of this nonparametric procedure are that it can be used with either ordinal data[1] or quantitative data and it does not require the assumption that the populations have a normal distribution. Versions of the test were developed jointly by Mann and Whitney and also by Wilcoxon. As a result, the test has been referred to as the *Mann-Whitney test* and the *Wilcoxon rank-sum test*. The tests are equivalent and both versions provide the same conclusion. In this section, we will refer to this nonparametric test as the **Mann-Whitney-Wilcoxon (MWW) test**.

---

[1]Ordinal data are categorical data that can be rank ordered. This scale of measurement was discussed more fully in Section 1.2 of Chapter 1.

We begin the MWW test by stating the most general form of the null and alternative hypotheses as follows:

$H_0$: The two populations are identical

$H_a$: The two populations are not identical

The alternative hypothesis that the two populations are not identical requires some clarification. If $H_0$ is rejected, we are using the test to conclude that the populations are not identical and that population 1 tends to provide either smaller or larger values than population 2. A situation where population 1 tends to provide smaller values than population 2 is shown in Figure 19.4. Note that it is not necessary that all values from population 1 be less than all values from population 2. However, the figure correctly shows, the conclusion that $H_a$ is true; the two populations are not identical and population 1 tends to provide smaller values than population 2. In a two-tailed test, we consider the alternative hypothesis that either population may provide the smaller or larger values. One-tailed versions of the test can be formulated with the alternative hypothesis that population 1 provides either the smaller or the larger values compared to population 2.

We will first illustrate the MWW test using small samples with rank-ordered data. This will give you an understanding of how the rank-sum statistic is computed and how it is used to determine whether to reject the null hypothesis that the two populations are identical. Later in the section, we will introduce a large-sample approximation based on the normal distribution that will simplify the calculations required by the MWW test.

Let us consider the on-the-job performance ratings for employees at a Showtime Cinemas 20-screen multiplex movie theater. During an employee performance review, the theater manager rated all 35 employees from best (rating 1) to worst (rating 35) in the theater's annual report. Knowing that the part-time employees were primarily college and high school students, the district manager asked if there was evidence of a significant difference in performance for college students compared to high school students. In terms of the population of college students and the population of high school students who could be considered for employment at the theater, the hypotheses were stated as follows:

$H_0$: College and high school student populations are identical in terms of performance

$H_a$: College and high school student populations are not identical in terms of performance

We will use a .05 level of significance for this test.

We begin by selecting a random sample of four college students and a random sample of five high school students working at Showtime Cinemas. The theater manager's overall performance rating based on all 35 employees was recorded for each of these employees, as shown in Table 19.8. The first college student selected was rated 15th in the manager's annual performance report, the second college student selected was rated 3rd in the manager's annual performance report, and so on.

**FIGURE 19.4**    TWO POPULATIONS ARE NOT IDENTICAL WITH POPULATION 1 TENDING TO PROVIDE THE SMALLER VALUES
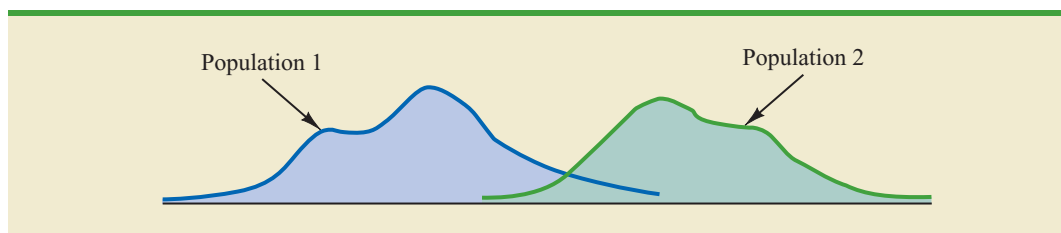
**TABLE 19.8**    PERFORMANCE RATINGS FOR A SAMPLE OF COLLEGE STUDENTS
AND A SAMPLE OF HIGH SCHOOL STUDENTS WORKING
AT SHOWTIME CINEMAS

| College Student | Manager's Performance Rating | High School Student | Manager's Performance Rating |
|---|---|---|---|
| 1 | 15 | 1 | 18 |
| 2 | 3 | 2 | 20 |
| 3 | 23 | 3 | 32 |
| 4 | 8 | 4 | 9 |
| | | 5 | 25 |

*The data in this example show how the MWW test can be used with ordinal (rank-ordered) data. Exercise 17 provides another application that uses this type of data.*

The next step in the MWW procedure is to rank the *combined* samples from low to high. Since there is a total of 9 students, we rank the performance rating data in Table 19.8 from 1 to 9. The lowest value of 3 for college student 2 receives a rank of 1 and the second lowest value of 8 for college student 4 receives a rank of 2. The highest value of 32 for high school student 3 receives a rank of 9. The combined-sample ranks for all 9 students are shown in Table 19.9.

Next we sum the ranks for each sample as shown in Table 19.9. The MWW procedure may use the sum of the ranks for either sample. However, in our application of the MWW test we will follow the common practice of using the first sample which is the sample of four college students. The sum of ranks for the first sample will be the test statistic $W$ for the MWW test. This sum, as shown in Table 19.9, is $W = 4 + 1 + 7 + 2 = 14$.

Let us consider why the sum of the ranks will help us select between the two hypotheses: $H_0$: The two populations are identical and $H_a$: The two populations are not identical. Letting C denote a college student and H denote a high school student, suppose the ranks of the nine students had the following order with the four college students having the four lowest ranks.

Rank     1   2   3   4   5   6   7   8   9
Student   C   C   C   C   H   H   H   H   H

Notice that this permutation or ordering separates the two samples, with the college students all having a lower rank than the high school students. This is a strong indication that the two populations are not identical. The sum of ranks for the college students in this case is $W = 1 + 2 + 3 + 4 = 10$.

**TABLE 19.9**    RANKS FOR THE NINE STUDENTS IN THE SHOWTIME CINEMAS
COMBINED SAMPLES

| College Student | Manager's Performance Rating | Rank | High School Student | Manager's Performance Rating | Rank |
|---|---|---|---|---|---|
| 1 | 15 | 4 | 1 | 18 | 5 |
| 2 | 3 | 1 | 2 | 20 | 6 |
| 3 | 23 | 7 | 3 | 32 | 9 |
| 4 | 8 | 2 | 4 | 9 | 3 |
| | Sum of Ranks | 14 | 5 | 25 | 8 |
| | | | | Sum of Ranks | 31 |

Now consider a ranking where the four college students have the four highest ranks.

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Student | H | H | H | H | H | C | C | C | C |

Notice that this permutation or ordering separates the two samples again, but this time the college students all have a higher rank than the high school students. This is another strong indication that the two populations are not identical. The sum of ranks for the college students in this case is $W = 6 + 7 + 8 + 9 = 30$. Thus, we see that the sum of the ranks for the college students must be between 10 and 30. Values of $W$ near 10 imply that college students have lower ranks than the high school students, whereas values of $W$ near 30 imply that college students have higher ranks than the high school students. Either of these extremes would signal the two populations are not identical. However, if the two populations are identical, we would expect a mix in the ordering of the C's and H's so that the sum of ranks $W$ is closer to the average of the two extremes, or nearer to $(10 + 30)/2 = 20$.

Making the assumption that the two populations are identical, we used a computer program to compute all possible orderings for the nine students. For each ordering, we computed the sum of the ranks for the college students. This provided the probability distribution showing the exact sampling distribution of $W$ in Figure 19.5. The exact probabilities associated with the values of $W$ are summarized in Table 19.10. While we will not ask you to generate this exact sampling distribution, we will use it to test the hypothesis that the two populations of students are identical.

Let us use the sampling distribution of $W$ in Figure 19.5 to compute the $p$-value for the test just as we have done using other sampling distributions. Table 19.9 shows that the sum of ranks for the four college student is $W = 14$. Because this value of $W$ is in the lower tail of the sampling distribution, we begin by computing the lower tail probability $P(W \le 14)$. Thus, we have

$$P(W \le 14) = P(10) + P(11) + P(12) + P(13) + P(14)$$
$$= .0079 + .0079 + .0159 + .0238 + .0397 = .0952$$

**FIGURE 19.5**   EXACT SAMPLING DISTRIBUTION OF THE SUM OF THE RANKS
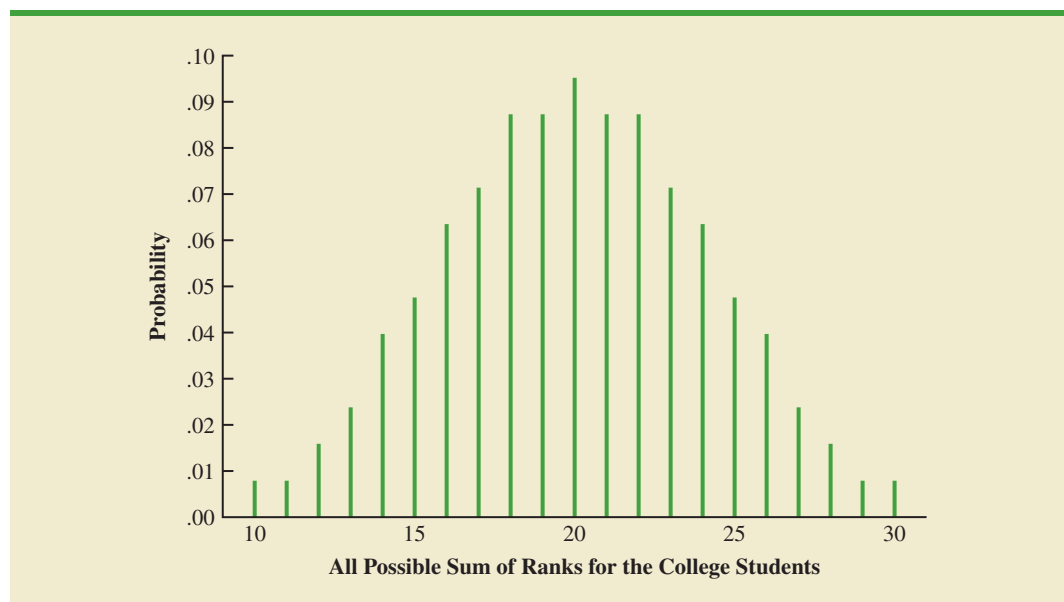FOR THE SAMPLE OF COLLEGE STUDENTS

**TABLE 19.10**   PROBABILITIES FOR THE EXACT SAMPLING DISTRIBUTION OF THE
SUM OF THE RANKS FOR THE SAMPLE OF COLLEGE STUDENTS

| W | Probability | W | Probability |
|---|---|---|---|
| 10 | 0.0079 | 20 | 0.0952 |
| 11 | 0.0079 | 21 | 0.0873 |
| 12 | 0.0159 | 22 | 0.0873 |
| 13 | 0.0238 | 23 | 0.0714 |
| 14 | 0.0397 | 24 | 0.0635 |
| 15 | 0.0476 | 25 | 0.0476 |
| 16 | 0.0635 | 26 | 0.0397 |
| 17 | 0.0714 | 27 | 0.0238 |
| 18 | 0.0873 | 28 | 0.0159 |
| 19 | 0.0873 | 29 | 0.0079 |
|  |  | 30 | 0.0079 |

The two-tailed $p$-value $= 2(.0952) = .1904$. With $\alpha = .05$ as the level of significance and $p$-value $> .05$, the MWW test conclusion is that we cannot reject the null hypothesis that the populations of college and high school students are identical. While the sample of four college students and the sample of five high school students did not provide statistical evidence to conclude there is a difference between the two populations, this is an ideal time to suggest withholding judgment. Further study with larger samples should be considered before drawing a final conclusion.

Most applications of the MWW test involve larger sample sizes than shown in this first example. For such applications, a large sample approximation of the sampling distribution of $W$ based on the normal distribution is employed. In fact, note that the sampling distribution of $W$ in Figure 19.5 shows a normal distribution is a pretty good approximation for sample sizes as small as four and five. We will use the same combined-sample ranking procedure that we used in the previous example but will use the normal distribution approximation rather than the exact sampling distribution of $W$ to compute the $p$-value and draw the conclusion.

We illustrate the use of the normal distribution approximation for the MWW test by considering the situation at Third National Bank. The bank manager is monitoring the balances maintained in checking accounts at two branch banks and is wondering if the populations of account balances at the two branch banks are identical. Two independent samples of checking accounts are taken with sample sizes $n_1 = 12$ at branch 1 and $n_2 = 10$ at branch 2. The data are shown in Table 19.11.

*Doing the ranking of the combined samples by hand will take some time. Computer routines can be used to do this ranking quickly and efficiently.*

As before, the first step in the MWW test is to rank the *combined* data from the lowest to highest values. Using the combined 22 observations in Table 19.11, we find the smallest value of $750 (Branch 2 Account 6) and assign it a rank of 1. The second smallest value of $800 (Branch 2 Account 5) is assigned a rank of 2. The third smallest value of 805 (Branch 1 Account 7) is assigned a rank of 3, and so on. In ranking the combined data, we may find that two or more values are the same. In that case, the tied values are assigned the *average* rank of their positions in the combined data set. For example, the balance of $950 occurs for both Branch 1 Account 6 and Branch 2 Account 4. In the combined data set, the two values of $950 are in positions 12 and 13 when the combined data are ranked from low to high. As a result, these two accounts are assigned the average rank $(12 + 13)/2 = 12.5$. Table 19.12 shows the assigned ranks for the combined samples.

**TABLE 19.11**   ACCOUNT BALANCES FOR TWO BRANCHES OF THIRD NATIONAL BANK

| Branch 1 | | Branch 2 | |
|---|---|---|---|
| **Account** | **Balance ($)** | **Account** | **Balance ($)** |
| 1 | 1095 | 1 | 885 |
| 2 | 955 | 2 | 850 |
| 3 | 1200 | 3 | 915 |
| 4 | 1195 | 4 | 950 |
| 5 | 925 | 5 | 800 |
| 6 | 950 | 6 | 750 |
| 7 | 805 | 7 | 865 |
| 8 | 945 | 8 | 1000 |
| 9 | 875 | 9 | 1050 |
| 10 | 1055 | 10 | 935 |
| 11 | 1025 | | |
| 12 | 975 | | |

**TABLE 19.12**   ASSIGNED RANKS FOR THE COMBINED ACCOUNT BALANCE SAMPLES

| Branch | Account | Balance | Rank |
|---|---|---|---|
| 2 | 6 | 750 | 1 |
| 2 | 5 | 800 | 2 |
| 1 | 7 | 805 | 3 |
| 2 | 2 | 850 | 4 |
| 2 | 7 | 865 | 5 |
| 1 | 9 | 875 | 6 |
| 2 | 1 | 885 | 7 |
| 2 | 3 | 915 | 8 |
| 1 | 5 | 925 | 9 |
| 2 | 10 | 935 | 10 |
| 1 | 8 | 945 | 11 |
| 1 | 6 | 950 | 12.5 |
| 2 | 4 | 950 | 12.5 |
| 1 | 2 | 955 | 14 |
| 1 | 12 | 975 | 15 |
| 2 | 8 | 1000 | 16 |
| 1 | 11 | 1025 | 17 |
| 2 | 9 | 1050 | 18 |
| 1 | 10 | 1055 | 19 |
| 1 | 1 | 1095 | 20 |
| 1 | 4 | 1195 | 21 |
| 1 | 3 | 1200 | 22 |

We now return to the two separate samples and show the ranks from Table 19.12 for each account balance. These results are provided in Table 19.13. The next step is to sum the ranks for each sample: 169.5 for sample 1 and 83.5 for sample 2 are shown. As stated

**TABLE 19.13**  COMBINED RANKING OF THE DATA IN THE TWO SAMPLES FROM THIRD NATIONAL BANK

| | Branch 1 | | | Branch 2 | |
|---|---|---|---|---|---|
| **Account** | **Balance ($)** | **Rank** | **Account** | **Balance ($)** | **Rank** |
| 1 | 1095 | 20 | 1 | 885 | 7 |
| 2 | 955 | 14 | 2 | 850 | 4 |
| 3 | 1200 | 22 | 3 | 915 | 8 |
| 4 | 1195 | 21 | 4 | 950 | 12.5 |
| 5 | 925 | 9 | 5 | 800 | 2 |
| 6 | 950 | 12.5 | 6 | 750 | 1 |
| 7 | 805 | 3 | 7 | 865 | 5 |
| 8 | 945 | 11 | 8 | 1000 | 16 |
| 9 | 875 | 6 | 9 | 1050 | 18 |
| 10 | 1055 | 19 | 10 | 935 | 10 |
| 11 | 1025 | 17 | | Sum of Ranks | 83.5 |
| 12 | 975 | 15 | | | |
| | Sum of Ranks | 169.5 | | | |

previously, we will always follow the procedure of using the sum of the ranks for sample 1 as the test statistic $W$. Thus, we have $W = 169.5$. When both samples sizes are 7 or more, a normal approximation of the sampling distribution of $W$ can be used. Under the assumption that the null hypothesis is true and the populations are identical, the sampling distribution of the test statistic $W$ is as follows.

---

SAMPLING DISTRIBUTION OF $W$ WITH IDENTICAL POPULATIONS

$$\text{Mean: } \mu_W = \tfrac{1}{2}n_1(n_1 + n_2 + 1) \tag{19.5}$$

$$\text{Standard deviation: } \sigma_W = \sqrt{\tfrac{1}{12} n_1 n_2 (n_1 + n_2 + 1)} \tag{19.6}$$

Distribution form: Approximately normal provided $n_1 \geq 7$ and $n_2 \geq 7$

---

Given the sample sizes $n_1 = 12$ and $n_2 = 10$, equations (19.5) and (19.6) provide the following mean and standard deviation for the sampling distribution:

$$\mu_W = \tfrac{1}{2}n_1(n_1 + n_2 + 1) = \tfrac{1}{2}(12)(12 + 10 + 1) = 138$$

$$\sigma_W = \sqrt{\tfrac{1}{12} n_1 n_2 (n_1 + n_2 + 1)} = \sqrt{\tfrac{1}{12} (12)(10)(12 + 10 + 1)} = 15.1658$$

Figure 19.6 shows the normal distribution used for the sampling distribution of $W$.

Let us proceed with the MWW test and use a .05 level of significance to draw a conclusion. Since the test statistic $W$ is discrete and the normal distribution is continuous, we will again use the continuity correction factor for the normal distribution approximation.

**FIGURE 19.6**   SAMPLING DISTRIBUTION OF $W$ FOR THE THIRD NATIONAL
BANK EXAMPLE
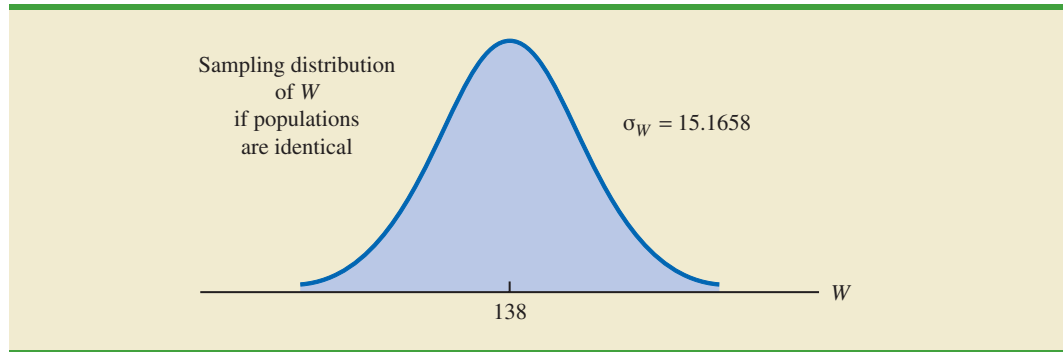


With $W = 169.5$ in the upper tail of the sampling distribution, we have the following $p$-value calculation:

$$P(W \geq 169.5) = P\left(z \geq \frac{169 - 138}{15.1658}\right) = P(z \geq 2.04)$$

Using the standard normal random variable and $z = 2.04$, the two-tailed $p$-value $= 2(1 - .9793) = .0414$. With $p$-value $\leq .05$, reject $H_0$ and conclude that the two populations of account balances are not identical. The upper tail value for test statistic $W$ indicates that the population of account balances at branch 1 tends to be larger.

*If the assumption can be made that the two populations have the same shape, the MWW test becomes a test about the difference between the medians of the two populations.*

As a final comment, some applications of the MWW test make it appropriate to assume that the two populations have identical shapes and if the populations differ, it is only by a shift in the location of the distributions. If the two populations have the same shape, the hypothesis test may be stated in terms of the difference between the two population medians. Any difference between the medians can be interpreted as the shift in location of one population compared to the other. In this case, the three forms of the MWW test about the medians of the two populations are as follows:

| **Two-Tailed Test** | **Lower Tail Test** | **Upper Tail Test** |
|---|---|---|
| $H_0$: Median$_1$ − Median$_2$ = 0 | $H_0$: Median$_1$ − Median$_2$ ≥ 0 | $H_0$: Median$_1$ − Median$_2$ ≤ 0 |
| $H_a$: Median$_1$ − Median$_2$ ≠ 0 | $H_a$: Median$_1$ − Median$_2$ < 0 | $H_a$: Median$_1$ − Median$_2$ > 0 |

## NOTES AND COMMENTS

The Minitab procedure for the MWW test is described in Appendix 19.1. Minitab makes the assumption that the two populations have the same shape. As a result, Minitab describes the test results in terms of a test about the medians of the two populations. If you do not feel comfortable making the "same shape" assumption, Minitab results are still applicable. However, you need to interpret the results as a test of the null hypothesis that the two populations are identical.

## Exercises

## Applications

**SELF** test

18. Two fuel additives are being tested to determine their effect on gasoline mileage. Seven cars were tested with additive 1 and nine cars were tested with additive 2. The following data show the miles per gallon obtained with the two additives. Use $\alpha = .05$ and the MWW test to see whether there is a significant difference between gasoline mileage for the two additives.

| Additive 1 | Additive 2 |
|------------|------------|
| 17.3 | 18.7 |
| 18.4 | 17.8 |
| 19.1 | 21.3 |
| 16.7 | 21.0 |
| 18.2 | 22.1 |
| 18.6 | 18.7 |
| 17.5 | 19.8 |
|  | 20.7 |
|  | 20.2 |

**SELF** test

19. Samples of starting annual salaries for individuals entering the public accounting and financial planning professions follow. Annual salaries are shown in thousands of dollars.

**WEB** file

**AcctPlanners**

| Public Accountant | Financial Planner |
|-------------------|-------------------|
| 50.2 | 49.0 |
| 58.8 | 49.2 |
| 56.3 | 53.1 |
| 58.2 | 55.9 |
| 54.2 | 51.9 |
| 55.0 | 53.6 |
| 50.9 | 49.7 |
| 59.5 | 53.9 |
| 57.0 | 51.8 |
| 51.9 | 48.9 |

   a. Use a .05 level of significance and test the hypothesis that there is no difference between the starting annual salaries of public accountants and financial planners. What is your conclusion?
   b. What are the sample median annual salaries for the two professions?

20. The gap between the earnings of men and women with equal education is narrowing but has not closed. Sample data for seven men and seven women with bachelor's degrees are as follows. Data are shown in thousands of dollars.

| Men | Women |
|------|-------|
| 35.6 | 49.5 |
| 80.5 | 40.4 |
| 50.2 | 32.9 |
| 67.2 | 45.5 |
| 43.2 | 30.8 |
| 54.9 | 52.5 |
| 60.3 | 29.8 |

    a.  What is the median salary for men? For women?

    b.  Use $\alpha = .05$ and conduct the hypothesis test for identical population distributions. What is your conclusion?

21.  Unisys maintains a hurricane database which provides information on hurricanes in the Atlantic/Caribbean/Gulf of Mexico and the Eastern Pacific Ocean. Maximum wind speeds in knots for a sample of hurricanes over the past four hurricane seasons are shown (Unisys weather website, April 2009).

**WEB** file

**Hurricanes**

| Atlantic/Carribean/Gulf of Mexico | | Eastern Pacific Ocean | |
|------|------|------|------|
| **Storm** | **Max Wind Speed** | **Storm** | **Max Wind Speed** |
| Frances | 125 | Darby | 105 |
| Jeanne | 110 | Frank | 75 |
| Lisa | 65 | Isis | 65 |
| Emily | 135 | Hilary | 90 |
| Ophelia | 80 | Max | 70 |
| Rita | 150 | Bud | 110 |
| Wilma | 150 | Daniel | 130 |
| Ernesto | 65 | Sergio | 95 |
| Florence | 80 | Cosme | 65 |
| Helene | 105 | Flossie | 120 |
| Dean | 145 | Henriette | 75 |
| Karen | 60 | Ivo | 70 |

    Use $\alpha = .05$ and test to determine whether the distribution of hurricane wind speeds is the same for these two regions. What is your conclusion?

22.  Each year *BusinessWeek* publishes statistics on the world's 1000 largest companies. A company's price/earnings (P/E) ratio is the company's current stock price divided by the latest 12 months' earnings per share. The following table shows the P/E ratios for a sample of 10 Japanese companies and 12 U.S. companies. Is the difference between the P/E ratios for the two countries significant? Use the MWW test and $\alpha = .01$ to support your conclusion.

**WEB** file

**JapanUS**

| Japan | | United States | |
|------|------|------|------|
| **Company** | **P/E Ratio** | **Company** | **P/E Ratio** |
| Sumitomo Corp. | 153 | Gannet | 19 |
| Kinden | 21 | Motorola | 24 |
| Heiwa | 18 | Schlumberger | 24 |
| NCR Japan | 125 | Oracle Systems | 43 |
| | | | (*Continued*) |

| Japan | | United States | |
|---|---|---|---|
| **Company** | **P/E Ratio** | **Company** | **P/E Ratio** |
| Suzuki Motor | 31 | Gap | 22 |
| Fuji Bank | 213 | Winn-Dixie | 14 |
| Sumintomo Chemical | 64 | Ingersoll-Rand | 21 |
| Seibu Railway | 666 | American Electric | 14 |
| Shiseido | 33 | Hercules | 21 |
| Toho Gas | 68 | Times Mirror | 38 |
| | | WellPoint Health | 15 |
| | | Northern States Power | 14 |

23. Police records show the following numbers of daily crime reports for a sample of days during the winter months and a sample of days during the summer months. Use a .05 level of significance to determine whether there is a significant difference between the winter and summer months in terms of the number of crime reports. What is your conclusion?

**WEB** file

**PoliceRecords**

| Winter | Summer |
|---|---|
| 18 | 28 |
| 20 | 18 |
| 15 | 24 |
| 16 | 32 |
| 21 | 18 |
| 20 | 29 |
| 12 | 23 |
| 16 | 38 |
| 19 | 28 |
| 20 | 18 |

24. A certain brand of microwave oven was priced at 10 stores in Dallas and 13 stores in San Antonio. The data follow. Use a .05 level of significance and test whether prices for the microwave oven are the same in the two cities.

**WEB** file

**Microwave**

| Dallas | San Antonio |
|---|---|
| 445 | 460 |
| 489 | 451 |
| 405 | 435 |
| 485 | 479 |
| 439 | 475 |
| 449 | 445 |
| 436 | 429 |
| 420 | 434 |
| 430 | 410 |
| 405 | 422 |
| | 425 |
| | 459 |
| | 430 |

25. The National Football League (NFL) holds its annual draft of the nation's best college football players in April each year. Prior to the draft, various sporting news services project the players who will be drafted along with the order in which each will be selected. The better players are selected earlier in the draft. For the 2009 draft, the colleges from the Southeastern Conference (SEC) and the Atlantic Coast Conference (ACC) were projected to have the most players who would be selected during the first round (SportProjection website, March 15, 2009). The player's college and the projected draft position for seven players from each conference are as follows.

| Southeastern Conference | | Atlantic Coast Conference | |
|---|---|---|---|
| **Player's College** | **Projected Draft Position** | **Player's College** | **Projected Draft Position** |
| Georgia | 1 | Georgia Tech | 3 |
| Alabama | 2 | Wake Forest | 6 |
| Vanderbilt | 14 | Virginia | 8 |
| Florida | 18 | Wake Forest | 23 |
| Mississippi | 20 | Florida State | 25 |
| Mississippi | 24 | Maryland | 26 |
| Auburn | 27 | Virginia | 29 |

Using the projected draft position as an indicator of preference the NFL teams have for the two conferences, use the MWW test to determine if there is any difference between the NFL preferences for players from these two conferences? Use $\alpha = .05$. What is the $p$-value? What is your conclusion?

## (19.4) Kruskal–Wallis Test

In this section we extend the nonparametric procedures to hypothesis tests involving three or more populations. We considered a parametric test for this situation in Chapter 13 when we used quantitative data and assumed that the populations had normal distributions with the same standard deviations. Based on an independent random sample from each population, we used the $F$ distribution to test for differences among the population means.

The nonparametric **Kruskal-Wallis test** is based on the analysis of independent random samples from each of $k$ populations. This procedure can be used with either ordinal data or quantitative data and does not require the assumption that the populations have normal distributions. The general form of the null and alternative hypotheses is as follows:

$$H_0\text{: All populations are identical}$$
$$H_a\text{: Not all populations are identical}$$

If $H_0$ is rejected, we will conclude that there is a difference among the populations with one or more populations tending to provide smaller or larger values compared to the other populations. We will demonstrate the Kruskal-Wallis test using the following example.

Williams Manufacturing Company hires employees for its management staff from three different colleges. Recently, the company's personnel director began reviewing the annual performance reports for the management staff in an attempt to determine whether there are differences in the performance ratings among the managers who graduated from the three colleges. Performance rating data are available for independent samples of seven managers who graduated from college A, six managers who graduated from college B, and seven managers who graduated from college C. These data are summarized in Table 19.14. The performance rating shown for each manger is recorded on a scale from 0 to 100, with 100 being the highest possible rating. Suppose we want to test whether the three populations of managers are identical in terms of performance ratings. We will use a .05 level of significance for the test.

The first step in the Kruskal-Wallis procedure is to rank the combined samples from lowest to highest values. Using all 20 observations in Table 19.14, the lowest rating of 15 for the 4th manager in the college B sample receives a rank of 1. The highest rating of 95 for the 5th manager in the college A sample receives a rank of 20. The performance rating data and their assigned ranks are shown in Table 19.15. Note that we assigned the average ranks to tied performance ratings of 60, 70, 80, and 90. Table 19.15 also shows the sum of ranks for each of the three samples.

The Kruskal-Wallis test statistic uses the sum of the ranks for the three samples and is computed as follows.

**TABLE 19.14**

PERFORMANCE EVALUATION RATINGS FOR 20 WILLIAMS EMPLOYEES

| College A | College B | College C |
|---|---|---|
| 25 | 60 | 50 |
| 70 | 20 | 70 |
| 60 | 30 | 60 |
| 85 | 15 | 80 |
| 95 | 40 | 90 |
| 90 | 35 | 70 |
| 80 |  | 75 |

KRUSKAL-WALLIS TEST STATISTIC

$$H = \left[ \frac{12}{n_T(n_T + 1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right] - 3(n_T + 1) \qquad \textbf{(19.7)}$$

where

$$k = \text{the number of populations}$$
$$n_i = \text{the number of observations in sample } i$$
$$n_T = \sum_{i=1}^{k} n_i = \text{the total number of observations in all samples}$$
$$R_i = \text{the sum of the ranks for sample } i$$

**TABLE 19.15**   COMBINED RANKINGS FOR THE THREE SAMPLES

| College A | Rank | College B | Rank | College C | Rank |
|---|---|---|---|---|---|
| 25 | 3 | 60 | 9 | 50 | 7 |
| 70 | 12 | 20 | 2 | 70 | 12 |
| 60 | 9 | 30 | 4 | 60 | 9 |
| 85 | 17 | 15 | 1 | 80 | 15.5 |
| 95 | 20 | 40 | 6 | 90 | 18.5 |
| 90 | 18.5 | 35 | 5 | 70 | 12 |
| 80 | 15.5 | Sum of Ranks | 27 | 75 | 14 |
| Sum of Ranks | 95 | | | Sum of Ranks | 88 |

Kruskal and Wallis were able to show that, under the null hypothesis assumption of identical populations, the sampling distribution of $H$ can be approximated by a chi-square distribution with $(k - 1)$ degrees of freedom. This approximation is generally acceptable if the sample sizes for each of the $k$ populations are all greater than or equal to five. The null hypothesis of identical populations will be rejected if the test statistic $H$ is large. As a result, the Kruskal-Wallis test is always expressed as an upper tail test. The computation of the test statistic for the sample data in Table 19.15 is as follows.

The sample sizes are

$$n_1 = 7 \qquad n_2 = 6 \qquad n_3 = 7$$

and

$$n_T = \sum_{i=1}^{3} n_t = 7 + 6 + 7 = 20$$

Using the sum of ranks for each sample, the value of the Kruskal-Wallis test statistic is as follows:

$$H = \left[ \frac{12}{n_T(n_T + 1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right] - 3(n_T + 1) = \frac{12}{20(21)} \left[ \frac{(95)^2}{7} + \frac{(27)^2}{6} + \frac{(88)^2}{7} \right] - 3(20 + 1) = 8.92$$

We can now use the chi-square distribution table (Table 3 of Appendix B) to determine the $p$-value for the test. Using $k - 1 = 3 - 1 = 2$ degrees of freedom, we find $\chi^2 = 7.378$ has an area of .025 in the upper tail of the chi-square distribution and $\chi^2 = 9.21$ has an area of .01 in the upper tail of the chi-square distribution. With $H = 8.92$ between 7.378 and 9.21, we can conclude that the area in the upper tail of the chi-square distribution is between .025 and .01. Because this is an upper tail test, we conclude that the $p$-value is between .025 and .01. Using Minitab or Excel will show the exact $p$-value for $\chi^2 = 8.92$ is .0116. Because $p$-value $\leq \alpha = .05$, we reject $H_0$ and conclude that the three populations are not all the same. The three populations of performance ratings are not identical and differ significantly depending upon the college. Because the sum of the ranks is relatively low for the sample of managers who graduated from college B, it would be reasonable for the company to either reduce its recruiting from college B, or at least evaluate the college B graduates more thoroughly before making a hiring decision.

*If the assumption can be made that the populations all have the same shape, the Kruskal-Wallis test becomes a test about the medians of the k populations.*

As a final comment, we note that in some applications of the Kruskal-Wallis test it may be appropriate to make the assumption that the populations have identical shapes and if they differ, it is only by a shift in location for one or more of the populations. If the $k$ populations are assumed to have the same shape, the hypothesis test can be stated in terms of the population medians. In this case, the hypotheses for the Kruskal-Wallis test would be written as follows:

$$H_0: \text{Median}_1 = \text{Median}_2 = \cdots = \text{Median}_k$$
$$H_a: \text{Not all Medians are equal}$$

## NOTES AND COMMENTS

1. The example in this section used quantitative data on employee performance ratings to conduct the Kruskal-Wallis test. This test could also have been used if the data were the ordinal rankings of the 20 employees in terms of performance. In this case, the test would use the ordinal data directly. The step of converting the quantitative data into rank-ordered data would not be necessary. Exercise 30 illustrates this situation.

**2.** The Minitab procedure for the Kruskal-Wallis test is described in Appendix 19.1. Minitab makes the assumption that the populations all have the same shape. As a result, Minitab describes the Kruskal-Wallis test as a test of differences among the population medians. If you do not feel comfortable making the "same shape" assumption, you can still use Minitab. However, you will need to interpret the results as a test of the null hypothesis that all populations are identical.

## Exercises

### Applications

**SELF** test

**26.** A sample of 15 consumers provided the following product ratings for three different products. Five consumers were randomly assigned to test and rate each product. Use the Kruskal-Wallis test and $\alpha = .05$ to determine whether there is a significant difference among the ratings for the products.

| | Product | |
|---|---|---|
| **A** | **B** | **C** |
| 50 | 80 | 60 |
| 62 | 95 | 45 |
| 75 | 98 | 30 |
| 48 | 87 | 58 |
| 65 | 90 | 57 |

**WEB** file

**TestPrepare**

**27.** Three admission test preparation programs are being evaluated. The scores obtained by a sample of 20 people who used the programs provided the following data. Use the Kruskal-Wallis test to determine whether there is a significant difference among the three test preparation programs. Use $\alpha = .05$.

| | Program | |
|---|---|---|
| **A** | **B** | **C** |
| 540 | 450 | 600 |
| 400 | 540 | 630 |
| 490 | 400 | 580 |
| 530 | 410 | 490 |
| 490 | 480 | 590 |
| 610 | 370 | 620 |
| | 550 | 570 |

**SELF** test

**28.** Forty-minute workouts of one of the following activities three days a week will lead to a loss of weight. The following sample data show the number of calories burned during 40-minute workouts for three different activities. Do these data indicate differences in the amount of calories burned for the three activities? Use a .05 level of significance. What is your conclusion?

| Swimming | Tennis | Cycling |
|----------|--------|---------|
| 408 | 415 | 385 |
| 380 | 485 | 250 |
| 425 | 450 | 295 |
| 400 | 420 | 402 |
| 427 | 530 | 268 |

**WEB** file

**CruiseShips**

29. *Condé Nast Traveler* magazine conducts an annual survey of its readers in order to rate the top 80 cruise ships in the world (*Condé Nast Traveler*, February 2008). With 100 the highest possible rating, the overall ratings for a sample of ships from the Holland America, Princess, and Royal Caribbean cruise lines are shown. Use the Kruskal-Wallis test with $\alpha = .05$ to determine whether the overall ratings among the three cruise lines differ significantly. What is your conclusion?

| **Holland America** | | **Princess** | | **Royal Caribbean** | |
|---------------------|--------|--------------|--------|---------------------|--------|
| **Ship** | **Rating** | **Ship** | **Rating** | **Ship** | **Rating** |
| Amsterdam | 84.5 | Coral | 85.1 | Adventure | 84.8 |
| Maasdam | 81.4 | Dawn | 79.0 | Jewel | 81.8 |
| Ooterdam | 84.0 | Island | 83.9 | Mariner | 84.0 |
| Volendam | 78.5 | Princess | 81.1 | Navigator | 85.9 |
| Westerdam | 80.9 | Star | 83.7 | Serenade | 87.4 |

30. A large corporation sends many of its first-level managers to an off-site supervisory skills training course. Four different management development centers offer this course. The director of human resources would like to know whether there is a difference among the quality of training provided at the four centers. An independent random sample of five employees was chosen from each training center. The employees were then ranked 1 to 20 in terms of supervisory skills. A rank of 1 was assigned to the employee with the best supervisory skills. The ranks are shown. Use $\alpha = .05$ and test whether there is a significant difference among the quality of training provided by the four programs.

| | **Course** | | |
|---|---|---|---|
| **A** | **B** | **C** | **D** |
| 3 | 2 | 19 | 20 |
| 14 | 7 | 16 | 4 |
| 10 | 1 | 9 | 15 |
| 12 | 5 | 18 | 6 |
| 13 | 11 | 17 | 8 |

31. The better-selling candies are often high in calories. Assume that the following data show the calorie content from samples of M&M's, Kit Kat, and Milky Way II. Test for significant differences among the calorie content of these three candies. At a .05 level of significance, what is your conclusion?

| M&M's | Kit Kat | Milky Way II |
|-------|---------|--------------|
| 230 | 225 | 200 |
| 210 | 205 | 208 |
| 240 | 245 | 202 |
| 250 | 235 | 190 |
| 230 | 220 | 180 |

# 19.5  Rank Correlation

The Pearson product moment correlation coefficient introduced in Chapter 3 is a measure of the linear association between two variables using quantitative data. In this section, we provide a correlation measure of association between two variables when ordinal or rank-ordered data are available. The **Spearman rank-correlation coefficient** has been developed for this purpose.

SPEARMAN RANK-CORRELATION COEFFICIENT

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 + 1)} \qquad (19.8)$$

where

$n$ = the number of observations in the sample
$x_i$ = the rank of observation $i$ with respect to the first variable
$y_i$ = the rank of observation $i$ with respect to the second variable
$d_i = x_i - y_i$

Let us illustrate the use of the Spearman rank-correlation coefficient. A company wants to determine whether individuals who had a greater potential at the time of employment turn out to have higher sales records. To investigate, the personnel director reviewed the original job interview reports, academic records, and letters of recommendation for 10 current members of the sales force. After the review, the director ranked the 10 individuals in terms of their potential for success at the time of employment and assigned the individual who had the most potential the rank of 1. Data were then collected on the actual sales for each individual during their first two years of employment. On the basis of the actual sales records, a second ranking of the 10 individuals based on sales performance was obtained. Table 19.16 provides the ranks based on potential as well as the ranks based on the actual performance.

Let us compute the Spearman rank-correlation coefficient for the data in Table 19.16. The computations are summarized in Table 19.17. We first compute the difference between the two ranks for each salesperson, $d_i$, as shown in column 4. The sum of the $d_i^2$ in column 5 is 44. This value and the sample size $n = 10$ are used to compute the rank-correlation coefficient $r_s = .733$ shown in Table 19.17.

The Spearman rank-correlation coefficient ranges from $-1.0$ to $+1.0$ and its interpretation is similar to the Pearson product moment correlation coefficient for quantitative data. A rank-correlation coefficient near $+1.0$ indicates a strong positive association between the

**TABLE 19.16** SALES POTENTIAL AND ACTUAL TWO-YEAR SALES DATA

| Salesperson | Ranking of Potential | Two-Year Sales (units) | Ranking According to Two-Year Sales |
|---|---|---|---|
| A | 2 | 400 | 1 |
| B | 4 | 360 | 3 |
| C | 7 | 300 | 5 |
| D | 1 | 295 | 6 |
| E | 6 | 280 | 7 |
| F | 3 | 350 | 4 |
| G | 10 | 200 | 10 |
| H | 9 | 260 | 8 |
| I | 8 | 220 | 9 |
| J | 5 | 385 | 2 |

**WEB** file

**PotentialActual**

**TABLE 19.17** COMPUTATION OF THE SPEARMAN RANK-CORRELATION COEFFICIENT FOR SALES POTENTIAL AND SALES PERFORMANCE

| Salesperson | $x_i$ = Ranking of Potential | $y_i$ = Ranking of Sales Performance | $d_i = x_i - y_i$ | $d_i^2$ |
|---|---|---|---|---|
| A | 2 | 1 | 1 | 1 |
| B | 4 | 3 | 1 | 1 |
| C | 7 | 5 | 2 | 4 |
| D | 1 | 6 | −5 | 25 |
| E | 6 | 7 | −1 | 1 |
| F | 3 | 4 | −1 | 1 |
| G | 10 | 10 | 0 | 0 |
| H | 9 | 8 | 1 | 1 |
| I | 8 | 9 | −1 | 1 |
| J | 5 | 2 | 3 | 9 |
| | | | | $\Sigma d_i^2 = 44$ |

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 + 1)} = 1 - \frac{6(44)}{10(100 - 1)} = .733$$

ranks for the two variables, while a rank-correlation coefficient near −1.0 indicates a strong negative association between the ranks for the two variables. A rank-correlation coefficient of 0 indicates no association between the ranks for the two variables. In the example, $r_s$ = .733 indicates a positive correlation between the ranks based on potential and the ranks based on sales performance. Individuals who ranked higher in potential at the time of employment tended to rank higher in two-year sales performance.

At this point, we may want to use the sample rank correlation $r_s$ to make an inference about the population rank correlation coefficient $\rho_s$. To do this, we test the following hypotheses:

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

Under the assumption that the null hypothesis is true and the population rank-correlation co-efficient is 0, the following sampling distribution of $r_s$ can be used to conduct the test.

---

SAMPLING DISTRIBUTION OF $r_s$

$$\text{Mean: } \mu_{r_s} = 0 \qquad\qquad \textbf{(19.9)}$$

$$\text{Standard deviation: } \sigma_{r_s} = \sqrt{\frac{1}{n-1}} \qquad\qquad \textbf{(19.10)}$$

Distribution form: Approximately normal provided $n \geq 10$

---

The sample rank-correlation coefficient for sales potential and sales performance is $r_s = .733$. Using equation (19.9), we have $\mu_{r_s} = 0$, and using equation (19.10), we have $\sigma_{r_s} = \sqrt{1/(10-1)} = .333$. With the sampling distribution of $r_s$ approximated by a normal distribution, the standard normal random variable $z$ becomes the test statistic with

$$z = \frac{r_s - \mu_{r_s}}{\sigma_{r_s}} = \frac{.733 - 0}{.333} = 2.20$$

Using the standard normal probability table and $z = 2.20$, we find the two-tailed $p$-value = $2(1 - .9861) = .0278$. With a .05 level of significance, $p$-value $\leq \alpha$. Thus, we reject the null hypothesis that the population rank-correlation coefficient is zero. The test result shows that there is a significant rank correlation between potential at the time of employment and actual sales performance.

## NOTES AND COMMENTS

The Spearman rank-correlation coefficient provides the same value that is obtained by using the Pearson product moment correlation coefficient procedure with the rank-ordered data. In Appendixes 19.1 and 19.2, we show how Minitab and Excel correlation tools for the Pearson product moment correlation coefficient can be used to compute the Spearman rank-correlation coefficient.

## Exercises

## Methods

32.   Consider the following set of rankings for a sample of 10 elements.

**SELF** test

| Element | x | y | Element | x | y |
|---------|----|----|---------|----|----|
| 1 | 10 | 8 | 6 | 2 | 7 |
| 2 | 6 | 4 | 7 | 8 | 6 |
| 3 | 7 | 10 | 8 | 5 | 3 |
| 4 | 3 | 2 | 9 | 1 | 1 |
| 5 | 4 | 5 | 10 | 9 | 9 |

a.   Compute the Spearman rank-correlation coefficient for the data.
b.   Use $\alpha = .05$ and test for significant rank correlation. What is your conclusion?

33. Consider the following two sets of rankings for six items.

| | **Case One** | | | **Case Two** | |
|---|---|---|---|---|---|
| **Item** | **First Ranking** | **Second Ranking** | **Item** | **First Ranking** | **Second Ranking** |
| A | 1 | 1 | A | 1 | 6 |
| B | 2 | 2 | B | 2 | 5 |
| C | 3 | 3 | C | 3 | 4 |
| D | 4 | 4 | D | 4 | 3 |
| E | 5 | 5 | E | 5 | 2 |
| F | 6 | 6 | F | 6 | 1 |

Note that in the first case the rankings are identical, whereas in the second case the rankings are exactly opposite. What value should you expect for the Spearman rank-correlation coefficient for each of these cases? Explain. Calculate the rank-correlation coefficient for each case.

## Applications

**SELF test**

34. The following data show the rankings of 11 states based on expenditure per student (ranked 1 highest to 11 lowest) and student-teacher ratio (ranked 1 lowest to 11 highest).

**WEB file**

**Student**

| **State** | **Expenditure per Student** | **Student-Teacher Ratio** |
|---|---|---|
| Arizona | 9 | 10 |
| Colorado | 5 | 8 |
| Florida | 4 | 6 |
| Idaho | 2 | 11 |
| Iowa | 6 | 4 |
| Louisiana | 11 | 3 |
| Massachusetts | 1 | 1 |
| Nebraska | 7 | 2 |
| North Dakota | 8 | 7 |
| South Dakota | 10 | 5 |
| Washington | 3 | 9 |

a. What is the rank correlation between expenditure per student and student-teacher ratio. Discuss.

b. At the $\alpha = .05$ level, does there appear to be a relationship between expenditure per student and student-teacher ratio?

35. A national study by Harris Interactive, Inc., evaluated the top technology companies and their reputations. The following shows how 10 technology companies ranked in reputation and how the companies ranked in percentage of respondents who said they would purchase the company's stock. A positive rank correlation is anticipated because it seems reasonable to expect that a company with a higher reputation would have the more desirable stock to purchase.

| **Company** | **Reputation** | **Stock Purchase** |
|---|---|---|
| Microsoft | 1 | 3 |
| Intel | 2 | 4 |
| Dell | 3 | 1 |
| | | *(Continued)* |

| Company | Reputation | Stock Purchase |
|---|---|---|
| Lucent | 4 | 2 |
| Texas Instruments | 5 | 9 |
| Cisco Systems | 6 | 5 |
| Hewlett-Packard | 7 | 10 |
| IBM | 8 | 6 |
| Motorola | 9 | 7 |
| Yahoo | 10 | 8 |

**WEB file**

**Techs**

a.   Compute the rank correlation between reputation and stock purchase.
b.   Test for a significant positive rank correlation. What is the $p$-value?
c.   At $\alpha = .05$, what is your conclusion?

36.   The rankings of a sample of professional golfers in both driving distance and putting are shown. What is the rank correlation between driving distance and putting for these golfers? Test for significance of the correlation coefficient at the .10 level of significance.

| Golfer | Driving Distance | Putting |
|---|---|---|
| Fred Couples | 1 | 5 |
| David Duval | 5 | 6 |
| Ernie Els | 4 | 10 |
| Nick Faldo | 9 | 2 |
| Tom Lehman | 6 | 7 |
| Justin Leonard | 10 | 3 |
| Davis Love III | 2 | 8 |
| Phil Mickelson | 3 | 9 |
| Greg Norman | 7 | 4 |
| Mark O'Meara | 8 | 1 |

**WEB file**

**ProGolfers**

37.   A student organization surveyed both current students and recent graduates to obtain information on the quality of teaching at a particular university. An analysis of the responses provided the following teaching-ability rankings. Do the rankings given by the current students agree with the rankings given by the recent graduates? Use $\alpha = .10$ and test for a significant rank correlation.

| Professor | Current Students | Recent Graduates |
|---|---|---|
| 1 | 4 | 6 |
| 2 | 6 | 8 |
| 3 | 8 | 5 |
| 4 | 3 | 1 |
| 5 | 1 | 2 |
| 6 | 2 | 3 |
| 7 | 5 | 7 |
| 8 | 10 | 9 |
| 9 | 7 | 4 |
| 10 | 9 | 10 |

**WEB file**

**Professors**

## Summary

In this chapter we have presented statistical procedures that are classified as nonparametric methods. Because these methods can be applied to categorical data as well as quantitative data and because they do not require an assumption about the distribution of the population, they expand the number of situations that can be subjected to statistical analysis.

The sign test is a nonparametric procedure for testing a hypothesis about a population median or for testing a hypothesis with matched samples. The data must be summarized in two categories, one denoted by a plus sign and one denoted by a minus sign. The Wilcoxon signed-rank test analyzes matched samples from two populations when quantitative data are available. No assumption is required other than the distribution of the paired differences is symmetric. The Wilcoxon signed-rank test is used to determine if the median of the population of paired differences is zero. This test can also be used to make inferences about the median of a symmetric population.

The Mann-Whitney-Wilcoxon test is a nonparametric procedure for the difference between two populations based on two independent samples. It is an alternative to the parametric *t* test for the difference between the means of the two populations. The combined ranks for the data from the two samples are obtained and the test statistic for the MWW test is the sum of ranks for the first sample. In most applications, the samples sizes are large enough to use a normal approximation with the continuity correction factor in conducting the hypothesis test. If no assumption is made about the populations, the MWW procedure tests whether the two populations are identical. If the assumption can be made that the two populations have the same shape, the test provides an inference about the difference between the medians of the two populations.

The Kruskal-Wallis test extends the MWW test to three or more populations. It is an alternative to the parametric analysis of variance test for the differences among the means of three or more normally distributed populations. The Kruskal-Wallis test does not require any assumption about the distribution of the populations and uses the null hypothesis that the *k* populations are identical. If the assumption can be made that the populations have the same shape, the test provides an inference about differences among the medians of the *k* populations. In the last section of the chapter we introduced the Spearman rank-correlation coefficient as a measure of association between two variables based on rank-ordered data.

## Glossary

**Parametric methods** Statistical methods that begin with an assumption about the probability distribution of the population which is often that the population has a normal distribution. A sampling distribution for the test statistic can then be derived and used to make an inference about one or more parameters of the population such as the population mean $\mu$ or the population standard deviation $\alpha$.

**Nonparametric methods** Statistical methods that require no assumption about the form of the probability distribution of the population and are often referred to as distribution-free methods. Several of the methods can be applied with categorical as well as quantitative data.

**Distribution-free methods** Statistical methods that make no assumption about the probability distribution of the population.

**Sign test** A nonparametric test for a hypothesis about a population median or for identifying differences between two populations based on matched samples. The data are summarized in two categories, denoted by a plus sign or a minus sign, and the binomial distribution with $p = .50$ provides the sampling distribution for the test statistic.

**Wilcoxon signed-rank test** A nonparametric test for the difference between the medians of two populations based on matched samples. The procedure uses quantitative data and is based on the assumption that the distribution of differences is symmetric. The paired-difference data are used to make an inference about the medians of the two populations. This test can also be used to make inferences about the median of a symmetric population.

**Mann-Whitney-Wilcoxon (MWW) test** A nonparametric test for the difference between two populations based on an independent sample from each population. The null hypothesis is that the two populations are identical. If the assumption can be made that the populations have the same shape, this test provides an inference about the difference between the medians of the two populations.

**Kruskal-Wallis test** A nonparametric test for the differences among three or more populations based on the analysis of an independent sample from each population. The null hypothesis is that the populations are identical. If the assumption can be made that the populations have the same shape, this test provides an inference about the differences among the medians of the populations.

**Spearman rank-correlation coefficient** A correlation measure of the association between two variables based on rank-ordered data.

## Key Formulas

### Sign Test: Normal Approximation

$$\text{Mean: } \mu = .50n \tag{19.1}$$

$$\text{Standard Deviation: } \sigma = \sqrt{.25n} \tag{19.2}$$

### Wilcoxon Signed-Rank Test: Normal Approximation

$$\text{Mean: } \mu_{T^+} = \frac{n(n+1)}{4} \tag{19.3}$$

$$\text{Standard deviation: } \sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{19.4}$$

### Mann-Whitney-Wilcoxon Test: Normal Approximation

$$\text{Mean: } \mu_W = \tfrac{1}{2}n_1(n_1 + n_2 + 1) \tag{19.5}$$

$$\text{Standard deviation: } \sigma_W = \sqrt{\tfrac{1}{12}\, n_1 n_2 (n_1 + n_2 + 1)} \tag{19.6}$$

### Kruskal-Wallis Test Statistic

$$H = \left[ \frac{12}{n_T(n_T + 1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \right] - 3(n_T + 1) \tag{19.7}$$

### Spearman Rank-Correlation Coefficient

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 + 1)} \tag{19.8}$$

## Supplementary Exercises

38. A survey asked the following question: Do you favor or oppose providing tax-funded vouchers or tax deductions to parents who send their children to private schools? Of the 2010 individuals surveyed, 905 favored the proposal, 1045 opposed the proposal, and 60 offered no opinion. Do the data indicate a significant difference in the preferences for the financial support of parents who send their children to private schools? Use a .05 level of significance.

39. Due to a recent decline in the housing market, the national median sales price for single-family homes is $180,000 (The National Association of Realtors, January 2009). Assume that the following data were obtained from samples of recent sales of single-family homes in St. Louis and Denver.

| Metropolitan Area | Less Than $180,000 | Equal to $180,000 | Greater Than $180,000 |
|---|---|---|---|
| St. Louis | 32 | 2 | 18 |
| Denver | 13 | 1 | 27 |

a.  Is the median sales price in St. Louis significantly lower than the national median of $180,000? Use a statistical test with $\alpha = .05$ to support your conclusion.
b.  Is the median sales price in Denver significantly higher than the national median of $180,000? Use a statistical test with $\alpha = .05$ to support your conclusion.

40. Twelve homemakers were asked to estimate the retail selling price of two models of refrigerators. Their estimates of selling price are shown in the following table. Use these data and test at the .05 level of significance to determine whether there is a difference between the two models in terms of homemakers' perceptions of selling price.

**WEB** file

**Refrigerators**

| Homemaker | Model 1 | Model 2 | Homemaker | Model 1 | Model 2 |
|---|---|---|---|---|---|
| 1 | $850 | $1100 | 7 | $900 | $1090 |
| 2 | 960 | 920 | 8 | 890 | 1120 |
| 3 | 940 | 890 | 9 | 1100 | 1200 |
| 4 | 900 | 1050 | 10 | 700 | 890 |
| 5 | 790 | 1120 | 11 | 810 | 900 |
| 6 | 820 | 1000 | 12 | 920 | 900 |

41. A study was designed to evaluate the weight-gain potential of a new poultry feed. A sample of 12 chickens was used in a six-week study. The weight of each chicken was recorded before and after the six-week test period. The differences between the before and after weights of the 12 chickens are as follows: 1.5, 1.2, −.2, .0, .5, .7, .8, 1.0, .0, .6, .2,−.01. A positive difference indicates a weight gain and a negative difference indicates a weight loss. Use a .05 level of significance to determine if the new feed provides a significant weight gain for the chickens.

42. The following data are product weights for the same items produced on two different produc- tion lines. Test for a difference between the product weights for the two lines. Use $\alpha = .05$.

**WEB** file

**ProductWeights**

| Line 1 | Line 2 |
|---|---|
| 13.6 | 13.7 |
| 13.8 | 14.1 |
| 14.0 | 14.2 |
| 13.9 | 14.0 |
| 13.4 | 14.6 |
| 13.2 | 13.5 |
| 13.3 | 14.4 |
| 13.6 | 14.8 |
| 12.9 | 14.5 |
| 14.4 | 14.3 |
|  | 15.0 |
|  | 14.9 |

43. A client wants to determine whether there is a significant difference in the time required to complete a program evaluation with the three different methods that are in common use. The times (in hours) required for each of 18 evaluators to conduct a program evaluation follow. Use $\alpha = .05$ and test to see whether there is a significant difference in the time required by the three methods.

| Method 1 | Method 2 | Method 3 |
|----------|----------|----------|
| 68 | 62 | 58 |
| 74 | 73 | 67 |
| 65 | 75 | 69 |
| 76 | 68 | 57 |
| 77 | 72 | 59 |
| 72 | 70 | 62 |

WEB file

Methods

44. A sample of 20 engineers employed with a company for three years has been rank ordered with respect to managerial potential. Some of the engineers attended the company's management-development course, others attended an off-site management-development program at a local university, and the remainder did not attend any program. Use the following rankings and $\alpha = .025$ to test for a significant difference in the managerial potential of the three groups.

| No Program | Company Program | Off-Site Program |
|------------|-----------------|------------------|
| 16 | 12 | 7 |
| 9 | 20 | 1 |
| 10 | 17 | 4 |
| 15 | 19 | 2 |
| 11 | 6 | 3 |
| 13 | 18 | 8 |
|  | 14 | 5 |

WEB file

Programs

45. Course evaluation ratings for four college instructors are shown in the following table. Use $\alpha = .05$ and test for a significant difference among the rating for these instructors. What is your conclusion?

| | Instructor | | |
|-------|----------|---------|--------|
| Black | Jennings | Swanson | Wilson |
| 88 | 87 | 88 | 80 |
| 80 | 78 | 76 | 85 |
| 79 | 82 | 68 | 56 |
| 68 | 85 | 82 | 71 |
| 96 | 99 | 85 | 89 |
| 69 | 99 | 82 | 87 |
|  | 85 | 84 |  |
|  | 94 | 83 |  |
|  |  | 81 |  |

WEB file

Evaluations

46. A sample of 15 students received the following rankings on midterm and final examinations in a statistics course. Compute the Spearman rank-correlation coefficient for the data and test for a significant correlation with $\alpha = .10$. What is the $p$-value and what is your conclusion?

**WEB** file

**Exams**

| Rank | | Rank | | Rank | |
|---|---|---|---|---|---|
| Midterm | Final | Midterm | Final | Midterm | Final |
| 1 | 4 | 6 | 2 | 11 | 14 |
| 2 | 7 | 7 | 5 | 12 | 15 |
| 3 | 1 | 8 | 12 | 13 | 11 |
| 4 | 3 | 9 | 6 | 14 | 10 |
| 5 | 8 | 10 | 9 | 15 | 13 |

47. Nielsen Research provides weekly ratings of nationally broadcast television programs. The ratings of the 84 prime-time programs broadcast by the four major televisions networks (ABC, CBS, FOX, and NBC) for the week of April 14–20, 2008, are provided in the file named NielsenResearch. The ratings range from 1 to 103. Shown are the ratings for 12 shows in the file (days and times for shows that aired multiple episodes are shown). Do these data suggest that the overall ratings for the four networks differ significantly? Use the Kruskal-Wallis test with a .10 level of significance. What is the $p$-value and what is your conclusion?

**WEB** file

**NielsenResearch**

| Program | Network | Rating |
|---|---|---|
| *20/20* | ABC | 60 |
| *30 Rock* | NBC | 44 |
| *48 Hours Mystery* (Sat. 10:00 P.M) | CBS | 51 |
| *48 Hours Mystery* (Sat. 9:00 P.M) | CBS | 78 |
| *48 Hours Mystery* (Tues. 10:00 P.M) | CBS | 63 |
| *60 Minutes* | CBS | 13 |
| *According to Jim* (Tues. 8:00 P.M) | ABC | 89 |
| *According to Jim* (Tues. 8:30 P.M) | ABC | 91 |
| *American Dad* (Sun. 7:30 P.M) | FOX | 100 |
| *American Dad* (Sun. 9:30 P.M) | FOX | 65 |
| *American Idol* (Tues. 8:00 P.M) | FOX | 1 |
| *American Idol* (Wed. 9:00 P.M) | FOX | 2 |

## Appendix 19.1   Nonparametric Methods with Minitab

Minitab can be used for all the nonparametric methods introduced in this chapter.

### Sign Test for a Hypothesis Test About a Population Median

**WEB** file

**HomeSales**

We illustrate a hypothesis test about a population median using the sales price data for new homes in Section 19.1. The prices appear in column C1 of the Minitab worksheet named Home-Sales. The following steps can be used to test the hypotheses $H_0$: Median $\geq$ \$236,000 versus $H_a$: Median $<$ \$236,000.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Nonparametrics**
**Step 3.** Choose **1-Sample Sign**
**Step 4.** When the 1-Sample Sign dialogue box appears:
        Enter C1 in the **Variables** box
        Select **Test Median**
        Enter the hypothesized value 236000 in the **Test Median** box
        Select **less than** from the **Alternative** menu
        Click **OK**

Minitab provides the *p*-value as well as a point estimate of the population median.

This Minitab procedure can also be used to obtain an interval estimate of the population median. In step 4, select **Confidence interval** instead of **Test median**, enter the **Confidence level**, and click **OK**. For sample sizes greater than 50, Minitab uses a normal approximation to the binomial sampling distribution with the continuity correction factor for both the hypothesis test and the confidence interval calculations.

## Sign Test for a Hypothesis Test with Matched Samples

In order to use Minitab's sign test procedure for a hypothesis test with matched samples, we will use a numerical code for a plus sign, a minus sign, and the no preference data. We will use the Sun Coast Farms hypothesis test in Section 19.1 to illustrate this procedure. The data file SunCoast shows that column C1 contains numbers identifying each of the 14 individuals participating in the taste test and that column C2 uses a $+1$ for each plus sign, a $-1$ for each minus sign and a 0 for each no preference. If the null hypothesis of no preference is true, the median of the population of $+1$'s, $-1$'s, and 0's will be zero. Thus, we follow the steps for testing a population median with the median hypothesized to be zero. For the Sun Coast Farms hypothesis test we will use a two-tailed test as follows.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Nonparametrics**
**Step 3.** Choose **1-Sample Sign**
**Step 4.** When the 1-Sample Sign dialogue box appears:
        Enter C2 in the **Variables** box
        Select **Test Median**
        Enter the hypothesized value 0 in the **Test Median** box
        Select **not equal** from the **Alternative** menu
        Click **OK**

## Wilcoxon Signed-Rank Test with Matched Samples

The following steps can be used to test hypotheses about the difference between two population medians using matched-sample data. We will use the production task completion time data in Section 19.2 to illustrate. The data file MatchedSample provides the production times for method A in column C1, the production times for method B in column C2, and the differences in column C3. The following steps can be used to test the hypotheses $H_0$: Median $= 0$ and $H_a$: Median $\neq 0$ for the population of differences.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Nonparametrics**
**Step 3.** Choose **1-Sample Wilcoxon**

**Step 4.** When the 1-Sample Wilcoxon dialogue box appears:

> Enter C3 in the **Variables** box
> Select **Test Median**
> Enter the hypothesized value 0 in the **Test Median** box
> Select **not equal** from the **Alternative** menu
> Click **OK**

Note that the Minitab procedure uses the paired difference data in column C3. Although the data file shows the times for each production method in columns C1 and C2, these data were not used to obtain the Minitab output.

The same procedure can also be used to test a hypothesis about the median of a *symmetric* population. Enter the actual data in any column of the worksheet and follow the preceding steps. Enter the hypothesized value of the population median in the **Test Median** box and select the desired alternative hypothesis in the **Alternative** box. Click **OK** to obtain the results. For this test, you do not have to enter the difference data. The Minitab routine will make the calculations automatically. But remember, this test is valid for only the median of a symmetric population.

## Mann-Whitney-Wilcoxon Test

**WEB** file

**ThirdNational**

The following steps can be used to test a hypothesis that two populations are identical using two independent samples, one from each population. We will use the Third National Bank example in Section 19.3 to illustrate. The data file ThirdNational provides the 12 account balances for branch 1 in column C1 and the ten account balances for branch 2 in column C2. The following steps will implement the Minitab procedure for testing $H_0$: The two populations are identical versus $H_a$: The two populations are not identical.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Nonparametrics**
**Step 3.** Choose **Mann-Whitney**
**Step 4.** When the Mann-Whitney dialogue box appears:

> Enter C1 in the **First sample** box
> Enter C2 in the **Second sample** box
> Select **not equal** from the **Alternative** menu
> Click **OK**

Minitab will report the value of the test statistic and the corresponding *p*-value for the test. Since Minitab automatically assumes the two populations have the same shape, the output describes the results in terms of the difference between the medians of the two populations. Note that the output also provides both a point estimate and a confidence interval estimate of the difference between the medians. With the Greek letter $\eta$ (eta) sometimes used to denote a population median, the Minitab output uses ETA1 and ETA2 as abbreviations for the two population medians.

## Kruskal-Wallis Test

**WEB** file

**Williams**

The following steps can be used to test a hypothesis that three or more populations are identical using independent samples, one from each population. We will use the Williams Manufacturing Company data in Section 19.4 to illustrate. The data file Williams provides the college the employee attended (A, B, or C) in column C1 and the annual performance rating in column C2. Minitab's terminology is to refer to the college as the factor and the performance rating as the response. The following steps will implement the Minitab procedure for testing $H_0$: All populations are identical versus $H_a$: Not all populations are identical. If the assumption is made that the populations have the same shape, the hypotheses can be stated in terms of the population medians.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Nonparametrics**
**Step 3.** Choose **Kruskal-Wallis**
**Step 4.** When the Kruskal-Wallis dialogue box appears:
>Enter C2 in the **Response** box
>Enter C1 in the **Factor** box
>Click **OK**

## Spearman Rank Correlation

**WEB** file

**PotentialActual**

The Spearman rank-correlation coefficient is the same as the Pearson correlation coefficient computed for the ordinal, or rank-ordered, data. So we can compute the Spearman rank-correlation coefficient for the rank-ordered data by using Minitab's procedure to compute the Pearson correlation coefficient. We will use the sales potential and actual two-year sales data in Section 19.5 to illustrate. The data file PotentialActual provides the ranking of potential of each employee in column C2 and the ranking of the actual two-year sales of each employee in column C3. The following Minitab steps can be used to calculate Spearman's rank correlation for the two variables.

**Step 1.** Select the **Stat** menu
**Step 2.** Choose **Basic Statistics**
**Step 3.** Choose **Correlation**
**Step 4.** When the Correlation dialogue box appears:
>Enter C2 C3 in the **Variables** box
>Uncheck **Display p-values**
>Click **OK**

The Minitab output provides a value of .733 for the Pearson correlation coefficient. Since the data were rank-ordered data, this is also the Spearman rank-correlation coefficient. However, the $p$-value for the Pearson correlation coefficient is not appropriate for rank-ordered data and should not be interpreted as the $p$-value for the Spearman rank-correlation coefficient.

## Appendix 19.2 Nonparametric Methods with Excel

Excel does not have nonparametric procedures in its Data Analysis package. But we will show how Excel's BINOMDIST function can be used to conduct a sign test and how a Data Analysis procedure can be used to compute a rank-correlation coefficient. The StatTools Excel add-in can be used for the Wilcoxon signed-rank test and the Mann-Whitney-Wilcoxon test (see Appendix 19.3).

### Sign Test

The sign test uses a binomial sampling distribution with $p = .50$ to conduct a hypothesis test about a population median or a hypothesis test with matched samples. Excel's BINOMDIST function can be used to compute exact binomial probabilities for these tests. Since the BINOMDIST probabilities are exact, there is no need to use the normal distribution approximation calculation when using Excel for the sign test.

Let    $x =$ the number of plus signs

$n =$ the sample size for the observations with a plus sign or a minus sign

The BINOMDIST function can be used as follows:

$$\text{Lower tail probability} = \text{BINOMDIST}(x, n, .50, \text{True})$$
$$\text{Upper tail probability} = 1 - \text{BINOMDIST}(x - 1, n, .50, \text{True})$$

You can see from the lower tail probability expression, the BINOMDIST function provides the cumulative binomial probability of *less than or equal to x*. The .50 in the function is the value of $p = .50$ and the term True is used to obtain the cumulative binomial probability. The upper tail probability is $1 -$ (the cumulative probability) as shown. Note that since the binomial distribution is discrete, $(x - 1)$ is used in the upper tail probability calculation. For example, the upper tail probability $P(x \geq 7) = 1 - P(x \leq 6)$.

Using Excel for the Lawler Grocery Store hypothesis test about a population median, we have 7 plus signs and 3 minus signs for the sample of 10 stores. The number of plus signs was in the upper tail with $P(x \geq 7)$ given by the function

$$= 1 - \text{BINOMDIST}(x - 1, n, .50, \text{True}) = 1 - \text{BINOMDIST}(6, 10, .50, \text{True}) = .1719$$

Since this is a two-tailed hypothesis test, we have $p$-value $= 2(.1719) = .3438$.

In Section 19.1 we also considered the lower tail test about the population median price of a new home:

$$H_0: \text{Median} \geq 236{,}000$$
$$H_a: \text{Median} < 236{,}000$$

After deleting the home that sold for exactly \$236,000, the sample provided 22 plus signs and 38 minus signs for a sample of 60 homes. Since this is a lower tail test, the $p$-value is given by the lower tail probability $P(x \leq 22)$, which is as follows:

$$= \text{BINOMDIST}(x, n, .50, \text{True}) = \text{BINOMDIST}(22, 60, .50, \text{True}) = .0259$$

By using the BINOMDIST function, we have the capability of computing the exact $p$-value for any application of the sign test.

## Spearman Rank Correlation

**WEB** file

**PotentialActual**

Excel does not have a specific procedure for computing the Spearman rank-correlation coefficient. However, this correlation coefficient is the same as the Pearson correlation coefficient provided you are using rank-ordered data. As a result, we can compute the Spearman rank-correlation coefficient by applying Excel's Pearson correlation coefficient procedure to the rank-ordered data. We illustrate using the data on sales potential and actual two-year sales from Section 19.5. The data file PotentialActual provides the ranking of the 10 individuals in terms of potential in column B and the ranking of the 10 individuals in terms of actual two-year sales in column C. The following steps provide the Spearman rank-correlation coefficient.

**Step 1.** Click the **Data** tab on the Ribbon
**Step 2.** In the **Analysis** group, click **Data Analysis**
**Step 3.** Choose **Correlation** from the list of **Analysis Tools**
**Step 4.** When the Correlation Dialog box appears
　　　　Enter B1:C11 in the **Input Range** box
　　　　Select **Grouped by Columns**
　　　　Select **Labels in First Row**
　　　　Select **Output Range**
　　　　Enter D1 in the **Output Range** box
　　　　Click **OK**

The Spearman rank-correlation coefficient will appear in cell E3.

# Appendix 19.3    Nonparametric Methods with StatTools

In this appendix we show how to use StatTools for the Wilcoxon Signed-Rank test and the Mann-Whitney-Wilcoxon test.

## Wilcoxon Signed-Rank Test with Matched Samples

**WEB** file

**MatchedSample**

The following steps can be used to test hypotheses about the difference between two population medians based on matched samples. We will use the production task completion time data in Section 19.2 to illustrate. The data file MatchedSample provides the production times for method A in column A, the production times for method B in column B, and the differences between the two methods in column C. Begin by using the Data Set Manager to create a StatTools data set using the procedure described in the appendix to Chapter 1. The following steps can then be used to test the hypotheses $H_0$: Median $= 0$ and $H_a$: Median $\neq 0$ for the population of differences.

> **Step 1.** Click the S**tatTools** tab on the Ribbon
> **Step 2.** In the **Analyses Group**, select **Nonparametric Tests**
> **Step 3.** Choose **Wilcoxon Signed-Rank Test**
> **Step 4.** When the Wilcoxon Sign-Rank Test dialogue box appears:
>> Select **One-Sample Analysis** in the **Analysis Type** box
>> Check the **Difference** variable
>> Enter 0 in the **Null Hypothesis Value** box
>> Select **Not Equal to Null Value** in the **Alternative Hypothesis** box
>> Click **OK**

The same procedure can also be used to test a hypothesis about the median of a symmetric population. Enter the data in any column of the worksheet. Then follow the preceding steps. Enter the hypothesized value of the population median in the **Null Hypothesis Value** box and select the desired alternative hypothesis in the **Alternative Hypothesis** box. Click **OK** to obtain the results. For this test, you do not have to enter the difference data because the StatTools routine will make the calculations automatically. But remember, this test is valid only for the median of a symmetric population.

## Mann-Whitney-Wilcoxon Test

**WEB** file

**ThirdNational**

The following steps can be used to test a hypothesis that two populations are identical using two independent samples, one from each population. We will use the Third National Bank example in Section 19.3 to illustrate. The data file ThirdNational provides the 12 account balances for branch 1 in column A and the 10 account balances for branch 2 in column B. Begin by using the Data Set Manager to create a StatTools data set using the procedure described in the appendix to Chapter 1. The following steps can then be used for testing the hypotheses $H_0$: The two populations are identical and $H_a$: The two populations are not identical.

> **Step 1.** Click the **StatTools** tab on the Ribbon
> **Step 2.** In the **Analyses Group**, select **Nonparametric Tests**
> **Step 3.** Choose **Mann-Whitney Test**
> **Step 4.** When the Mann-Whitney Test dialogue box appears:
>> Select **General Version** in the **Analysis Type** box
>> Check the **Branch 1** variable
>> Check the **Branch 2** variable

Select **Either distribution smaller (Two-Tailed Test)** in the **Alternative Hypothesis** box

Click **OK**

**Step 5.** When the StatTools dialogue box appears:

Click **OK**

When the Choose Variable Ordering dialogue box appears:

Click **OK**

If you want to make the assumption that the two populations have the same shape, select **Median Version** in the **Analysis Type** box. The test results will be the same, with the output showing the hypotheses are about the difference between the two population medians.