# 11




Business Wire/Getty Images Publicity/Getty Images

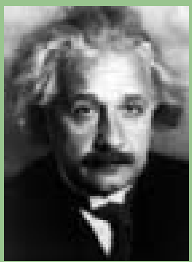


Photo Researchers / Alamy

*Make everything as simple as possible, but no simpler.*

—Albert Einstein

The brilliant German-born American physicist Albert Einstein (1879–1955) formulated the theory of relativity.

For online student resources, visit the Brase/Brase, *Understandable Statistics,* 10th edition web site at **http://www.cengage.com/statistics/brase**.

# Nonparametric Statistics

James Shafer/PhotoEdit

## PREVIEW QUESTIONS

*What if you cannot make assumptions about a population distribution? Can you still use statistical methods? What are the advantages and disadvantages?* (Section 11.1)

*What are nonparametric tests? How do you handle a "before and after" situation?* (Section 11.1)

*If you can't make assumptions about the population, and you have independent samples, how do you set up a nonparametric test?* (Section 11.2)

*Suppose you are interested only in rank data (ordinal-type data). If you have ordered pairs (x, y) of ranked data, is there a way to measure and test correlation?* (Section 11.3)

*Is a sequence random or is there a pattern associated with the sequence?* (Section 11.4)

## FOCUS PROBLEM

### How Cold? Compared to What?

Juneau is the capital of Alaska. The terrain surrounding Juneau is very rugged, and storms that sweep across the Gulf of Alaska usually hit Juneau. However, Juneau is located in southern Alaska, near the ocean, and temperatures are often comparable with those found in the lower 48 states. Madison is the capital of Wisconsin. The city is located between two large lakes. The climate of Madison is described as the typical continental climate of interior North America. Consider the long-term average temperatures (in degrees Fahrenheit) paired by month for the two cities (Source: National Weather Bureau). Use a sign test with a 5% level of significance to test the claim that the overall temperature distribution of Madison is different (either way) from that of Juneau. (See Problem 12 of Section 11.1.)

| Month | Madison | Juneau |
|---|---|---|
| January | 17.5 | 22.2 |
| February | 21.1 | 27.3 |
| March | 31.5 | 31.9 |
| April | 46.1 | 38.4 |
| May | 57.0 | 46.4 |
| June | 67.0 | 52.8 |
| July | 71.3 | 55.5 |
| August | 69.8 | 54.1 |
| September | 60.7 | 49.0 |
| October | 51.0 | 41.5 |
| November | 35.7 | 32.0 |
| December | 22.8 | 26.9 |

Medio Images / Photolibrary

SECTION 11.1

# The Sign Test for Matched Pairs

FOCUS POINTS

- State the criteria for setting up a matched pair sign test.
- Complete a matched pair sign test.
- Interpret the results in the context of the application.

Nonparametric statistics

There are many situations in which very little is known about the population from which samples are drawn. Therefore, we cannot make assumptions about the population distribution, such as assuming the distribution is normal or binomial. In this chapter, we will study methods that come under the heading of *nonparametric statistics*. These methods are called *nonparametric* because they require no assumptions about the population distributions from which samples are drawn. The obvious advantages of these tests are that they are quite general and (as we shall see) not difficult to apply. The disadvantages are that they tend to waste information and tend to result in acceptance of the null hypothesis more often than they should. As such, nonparametric tests are sometimes *less sensitive* than other tests.

Sign test

The easiest of all the nonparametric tests is probably the *sign test*. The sign test is used when we compare sample distributions from two populations that are *not independent*. This occurs when we measure the sample twice, as in "before and after" studies. The following example shows how the sign test is constructed and used:

Criteria for sign test

As part of their training, 15 police cadets took a special course on identification awareness. To determine how the course affects a cadet's ability to identify a suspect, the 15 cadets were first given an identification-awareness exam and then, after the course, were tested again. The police school would like to use the results of the two tests to see if the identification-awareness course *improves* a cadet's score. Table 11-1 gives the scores for each exam.

The sign of the difference is obtained by subtracting the precourse score from the postcourse score. If the difference is positive, we say that the sign of the difference is +, and if the difference is negative, we indicate it with −. No sign is indicated if the scores are identical; in essence, such scores are ignored when using the sign test. To use the sign test, we need to compute the *proportion x of plus signs* to all signs. We ignore the pairs with no difference of signs. This is demonstrated in Guided Exercise 1.

Photo Researchers, Inc.

**TABLE 11-1** **Scores for 15 Police Cadets**

| Cadet | Postcourse Score | Precourse Score | Sign of Difference |
|---|---|---|---|
| 1 | 93 | 76 | + |
| 2 | 70 | 72 | − |
| 3 | 81 | 75 | + |
| 4 | 65 | 68 | − |
| 5 | 79 | 65 | + |
| 6 | 54 | 54 | No difference |
| 7 | 94 | 88 | + |
| 8 | 91 | 81 | + |
| 9 | 77 | 65 | + |
| 10 | 65 | 57 | + |
| 11 | 95 | 86 | + |
| 12 | 89 | 87 | + |
| 13 | 78 | 78 | No difference |
| 14 | 80 | 77 | + |
| 15 | 76 | 76 | No difference |

**GUIDED EXERCISE 1** **Proportion of plus signs**

Look at Table 11-1 under the "Sign of Difference" column.

(a) How many plus signs do you see? → 10

(b) How many plus and minus signs do you see? → 12

(c) The *proportion of plus signs* is → $x = \frac{10}{12} = \frac{5}{6} \approx 0.833$

$$x = \frac{\text{Number of plus signs}}{\text{Total number of plus and minus signs}}$$

Use parts (a) and (b) to find $x$.

Null hypothesis

We observe that $x$ is the sample proportion of plus signs, and we use $p$ to represent the population proportion of plus signs (if *all* possible police cadets were tested). The null hypothesis is

$H_0$: $p = 0.5$ (the distributions of scores before and after the course are the same)

The null hypothesis states that the identification-awareness course does *not* affect the distribution of scores. Under the null hypothesis, we expect the number of plus signs and minus signs to be about equal. This means that the proportion of plus signs should be approximately 0.5.

Alternate hypothesis

The police department wants to see if the course *improves* a cadet's score. Therefore, the alternate hypothesis will be

$H_1$: $p > 0.5$ (the distribution of scores after the course is shifted higher than the distribution before the course)

The alternate hypothesis states that the identification-awareness course tends to improve scores. This means that the proportion of plus signs should be greater than 0.5.

Sampling distribution

To test the null hypothesis $H_0$: $p = 0.5$ against the alternate hypothesis $H_1$: $p > 0.5$, we use methods of Section 8.3 for tests of proportions. As in Section 8.3, we will assume that all our samples are sufficiently large to permit a normal approximation to the binomial distribution. For most practical work, this will be the case if the total number of plus and minus signs is 12 or more ($n \geq 12$).

When the total number of plus and minus signs is 12 or more, the sample statistic $x$ (proportion of plus signs) has a distribution that is approximately normal, with mean $p$ and standard deviation $\sqrt{pq/n}$ (See Section 6.6.)

Sample test statistic

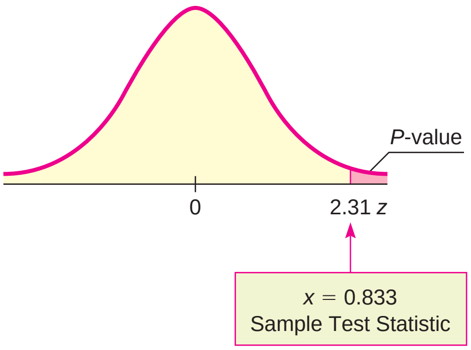
Under the null hypothesis $H_0$: $p = 0.5$, we assume that the population proportion $p$ of plus signs is 0.5. Therefore, the $z$ value corresponding to the sample test statistic $x$ is

$$z = \frac{x - p}{\sqrt{\frac{pq}{n}}} = \frac{x - 0.5}{\sqrt{\frac{(0.5)(0.5)}{n}}} = \frac{x - 0.5}{\sqrt{\frac{0.25}{n}}}$$

where $n$ is the total number of plus and minus signs, and $x$ is the total number of plus signs divided by $n$.

FIGURE 11-1

*P*-value




For the police cadet example, we found $x \approx 0.833$ in Guided Exercise 1. The value of $n$ is 12. (Note that of the 15 cadets in the sample, 3 had no difference in precourse and postcourse test scores, so there are no signs for these 3.) The $z$ value corresponding to $x = 0.833$ is then

$$z \approx \frac{0.833 - 0.5}{\sqrt{\frac{0.25}{12}}} \approx 2.31$$

*P*-value

We use the standard normal distribution table (Table 5 of Appendix II) to find *P*-values for the sign test. This table gives areas to the left of $z$. Recall from Section 8.2 that Table 5 of Appendix II can be used directly to find *P*-values of one-tailed tests. For *two-tailed* tests, we must *double* the value given in the table. To review the process of finding areas to the right or left of $z$ using Table 5, see Section 6.2.

The alternate hypothesis for the police cadet example is $H_1$: $p > 0.5$. The *P*-value for the sample test statistic $z = 2.31$ is shown in Figure 11-1. For a right-tailed test, the *P*-value is the area to the right of the sample test statistic $z = 2.31$. From Table 5 of Appendix II, $P(z > 2.31) = 0.0104$.

Conclude the test and interpret the results

In our example, the police department wishes to use a 5% level of significance to test the claim that the identification-awareness course improves a cadet's score. Since the *P*-value of 0.0104 is less than $\alpha = 0.05$, we reject the null hypothesis $H_0$ that the course makes no difference. Instead, at the 5% level of significance, we say the results are significant. The evidence is sufficient to claim that the identification-awareness course improves cadets' scores.

The steps used to construct a sign test for matched pairs are summarized in the next procedure.

**PROCEDURE**

## HOW TO CONSTRUCT A SIGN TEST FOR MATCHED PAIRS

***Setup and Requirements***

You first need a random sample of data pairs $(A, B)$. Next, you take the differences $A - B$ and record the sign change for each difference: plus, minus, or no change. The number of data pairs should be large enough that the total number of plus and minus signs is at least 12. The sample proportion of plus signs is

$$x = \frac{\text{number of plus signs}}{\text{total number of plus and minus signs}}$$

Let $p$ represent the population proportion of plus signs if the entire population of all possible data pairs $(A, B)$ were to be used.

*Continued*

***Procedure***

1. Set the *level of significance* $\alpha$. The *null hypothesis* is $H_0$: $p = 0.5$. In the context of the application, set the *alternate hypothesis:* $H_1$: $p > 0.5$, $H_1$: $p < 0.5$, or $H_1$: $p \neq 0.5$.
2. The *sample test statistic* is

$$z = \frac{x - 0.5}{\sqrt{\frac{0.25}{n}}}$$

where $n \geq 12$ is the total number of plus and minus signs.
3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the test statistic.
4. *Conclude* the test. If *P*-value $\leq \alpha$, then reject $H_0$. If *P*-value $> \alpha$, then do not reject $H_0$.
5. *Interpret your conclusion* in the context of the application.

## GUIDED EXERCISE 2 Sign test

Dr. Kick-a-poo's Traveling Circus made a stop at Middlebury, Vermont, where the doctor opened a booth and sold bottles of Dr. Kick-a-poo's Magic Gasoline Additive. The additive is supposed to increase gas mileage when used according to instructions. Twenty local people purchased bottles of the additive and used it according to instructions. These people carefully recorded their mileage with and without the additive. The results are shown in Table 11-2.

TABLE 11-2 **Mileage Before and After Kick-a-poo's Additive**

| Car | With Additive | Without Additive | Sign of Difference |
|---|---|---|---|
| 1 | 17.1 | 16.8 | + |
| 2 | 21.2 | 20.1 | + |
| 3 | 12.3 | 12.3 | No difference (N.D.) |
| 4 | 19.6 | 21.0 | − |
| 5 | 22.5 | 20.9 | + |
| 6 | 17.0 | 17.9 | ___ |
| 7 | 24.2 | 25.4 | ___ |
| 8 | 22.2 | 20.1 | ___ |
| 9 | 18.3 | 19.1 | ___ |
| 10 | 11.0 | 12.3 | ___ |
| 11 | 17.6 | 14.2 | ___ |
| 12 | 22.1 | 23.7 | ___ |
| 13 | 29.9 | 30.2 | ___ |
| 14 | 27.6 | 27.6 | ___ |
| 15 | 28.4 | 27.7 | ___ |
| 16 | 16.1 | 16.1 | ___ |
| 17 | 19.0 | 19.5 | ___ |
| 18 | 38.7 | 37.9 | ___ |
| 19 | 17.6 | 19.7 | ___ |
| 20 | 21.6 | 22.2 | ___ |

TABLE 11-3 **Completion of Table 11-2**

| Car | Sign of Difference |
|---|---|
| 6 | − |
| 7 | − |
| 8 | + |
| 9 | − |
| 10 | − |
| 11 | + |
| 12 | − |
| 13 | − |
| 14 | N.D. |
| 15 | + |
| 16 | N.D. |
| 17 | − |
| 18 | + |
| 19 | − |
| 20 | − |

*Continued*

GUIDED EXERCISE 2 *continued*

(a) In Table 11-2, complete the column headed "Sign of Difference." How many plus signs are there? How many total plus and minus signs are there? What is the value of $x$, the proportion of plus signs?

There are 7 plus signs and 17 total plus and minus signs. The proportion of plus signs is

$$x = \frac{7}{17} \approx 0.412$$

(b) Most people claim that the additive has no effect. Let's use a 0.05 level of significance to test this claim against the alternate hypothesis that the additive did have an effect (one way or the other). State the null and alternate hypotheses.

We use
$H_0$: $p = 0.5$ (mileage distributions are the same)
$H_1$: $p \neq 0.5$ (mileage distributions are different)

(c) Convert the sample $x$ value, $x = 0.412$, to a $z$ value.

To find the $z$ value corresponding to $x = 0.412$, we use $n = 17$ (total number of signs).

$$z = \frac{x - 0.5}{\sqrt{0.25/n}} \approx \frac{0.412 - 0.5}{\sqrt{0.25/17}} \approx -0.73$$

(d) Find the corresponding $P$-value.

FIGURE 11-2 *P*-value

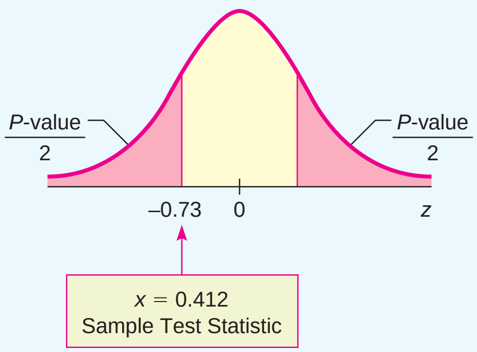



Table 5 of Appendix II gives the area to the left of $z = -0.73$.

$P(z < -0.73) = 0.2327$

Because this is a two-tailed test, the $P$-value is double this area.

$P\text{-value} = 2(0.2327) = 0.4654$

(e) Conclude the test.

For $\alpha = 0.05$, we see that the $P$-value $= 0.4654$ is greater than $\alpha$. We fail to reject $H_0$.

(f) ***Interpret*** the results.

At the 5% level of significance, the data are not statistically significant, and we cannot reject the hypothesis that the mileage distribution is the same with or without the additive.

## VIEWPOINT *Yukon News*

The Yukon News *featured an article entitled "Resurgence of the Dreaded White Plague," about the resurgence of tuberculosis (TB) in the far north. TB, also known as the white plague, has been present in Canada since it was brought in by European immigrants in the 17th century. Although antibiotics are widely used today, the disease has never been eradicated. Canadian National Health data suggest that TB is spreading faster in the Yukon than elsewhere in Canada. Because of this, the Canadian government has established many new TB clinics in remote Yukon villages. Using what you have learned in this section and Canadian National Health data, can you think of a way to use a sign test to study the claim that in these villages, the rate of TB in the population dropped after the clinics were activated?*

## SECTION 11.1 PROBLEMS

1. *Statistical Literacy* To apply the sign test, do you need independent or dependent (matched pair) data?

2. *Statistical Literacy* For the sign test of matched pairs, do pairs for which the difference in values is zero enter into any calculations?

For Problems 3–12, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Compute the sample test statistic. What is the sampling distribution?
(c) Find the $P$-value of the sample test statistic.
(d) Conclude the test.
(e) *Interpret* the conclusion in the context of the application.

3. *Economic Growth: Asia* Asian economies impact some of the world's largest populations. The growth of an economy has a big influence on the everyday lives of ordinary people. Are Asian economies changing? A random sample of 15 Asian economies gave the following information about annual percentage growth rate (Reference: *Handbook of International Economic Statistics,* U.S. Government Documents).

| Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Modern Growth Rate % | 4.0 | 2.3 | 7.8 | 2.8 | 0.7 | 5.1 | 2.9 | 4.2 |
| Historic Growth Rate % | 3.3 | 1.9 | 7.0 | 5.5 | 3.3 | 6.0 | 3.2 | 8.2 |

| Region | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|
| Modern Growth Rate % | 4.9 | 5.8 | 6.8 | 3.6 | 3.2 | 0.8 | 7.3 |
| Historic Growth Rate % | 6.4 | 7.2 | 6.1 | 1.5 | 1.0 | 2.1 | 5.1 |

Does this information indicate a change (either way) in the growth rate of Asian economies? Use a 5% level of significance.

4. *Debt: Developing Countries* Borrowing money may be necessary for business expansion. However, too much borrowed money can also mean trouble. Are developing countries tending to borrow more? A random sample of 20 developing countries gave the following information regarding foreign debt per capita (in U.S. dollars, inflation adjusted) (Reference: *Handbook of International Economic Statistics,* U.S. Government Documents).

| Country | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Modern Debt per Capita | 179 | 157 | 129 | 125 | 91 | 80 | 31 | 25 | 29 | 85 |
| Historic Debt per Capita | 144 | 132 | 88 | 112 | 53 | 66 | 31 | 30 | 40 | 75 |

| Country | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Modern Debt per Capita | 27 | 20 | 17 | 21 | 195 | 189 | 143 | 126 | 106 | 76 |
| Historic Debt per Capita | 21 | 19 | 15 | 24 | 104 | 150 | 142 | 118 | 117 | 79 |

Does this information indicate that foreign debt per capita is increasing in developing countries? Use a 1% level of significance.

5. *Education: Exams* A high school science teacher decided to give a series of lectures on current events. To determine if the lectures had any effect on student awareness of current events, an exam was given to the class before the lectures, and a similar exam was given after the lectures. The scores follow.

Use a 0.05 level of significance to test the claim that the lectures made no difference against the claim that the lectures did make some difference (one way or the other).

| Student | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| After Lectures | 107 | 115 | 120 | 78 | 83 | 56 | 71 | 89 | 77 |
| Before Lectures | 111 | 110 | 93 | 75 | 88 | 56 | 75 | 73 | 83 |

| Student | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|
| After Lectures | 44 | 119 | 130 | 91 | 99 | 96 | 83 | 100 | 118 |
| Before Lectures | 40 | 115 | 101 | 110 | 90 | 98 | 76 | 100 | 109 |

6. ***Grain Yields: Feeding the World*** With an ever-increasing world population, grain yields are extremely important. A random sample of 16 large grain-producing regions in the world gave the following information about grain production (in kg/hectare) (Reference: *Handbook of International Economic Statistics*, U.S. Government Documents).

| Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Modern Production | 1610 | 2230 | 5270 | 6990 | 2010 | 4560 | 780 | 6510 |
| Historic Production | 1590 | 2360 | 5161 | 7170 | 1920 | 4760 | 660 | 6320 |

| Region | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|
| Modern Production | 2850 | 3550 | 1710 | 2050 | 2750 | 2550 | 6750 | 3670 |
| Historic Production | 2920 | 2440 | 1340 | 2180 | 3110 | 2070 | 7330 | 2980 |

Does this information indicate that modern grain production is higher? Use a 5% level of significance.

7. ***Identical Twins: Reading Skills*** To compare two elementary schools regarding teaching of reading skills, 12 sets of identical twins were used. In each case, one child was selected at random and sent to school A, and his or her twin was sent to school B. Near the end of fifth grade, an achievement test was given to each child. The results follow:

| Twin Pair | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| School A | 177 | 150 | 112 | 95 | 120 | 117 |
| School B | 86 | 135 | 115 | 110 | 116 | 84 |

| Twin Pair | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|
| School A | 86 | 111 | 110 | 142 | 125 | 89 |
| School B | 93 | 77 | 96 | 130 | 147 | 101 |

Use a 0.05 level of significance to test the hypothesis that the two schools have the same effectiveness in teaching reading skills against the alternate hypothesis that the schools are not equally effective.

8. ***Incomes: Electricians and Carpenters*** How do the average weekly incomes of electricians and carpenters compare? A random sample of 17 regions in the United States gave the following information about average weekly income (in dollars) (Reference: U.S. Department of Labor, Bureau of Labor Statistics).

| Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Electricians | 461 | 713 | 593 | 468 | 730 | 690 | 740 | 572 | 805 |
| Carpenters | 540 | 812 | 512 | 473 | 686 | 507 | 785 | 657 | 475 |

| Region | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|
| Electricians | 593 | 593 | 700 | 572 | 863 | 599 | 596 | 653 |
| Carpenters | 485 | 646 | 675 | 382 | 819 | 600 | 559 | 501 |

Does this information indicate a difference (either way) in the average weekly incomes of electricians compared to those of carpenters? Use a 5% level of significance.

9. ***Quitting Smoking: Hypnosis*** One program to help people stop smoking cigarettes uses the method of posthypnotic suggestion to remind subjects to avoid smoking. A random sample of 18 subjects agreed to test the program. All subjects counted the number of cigarettes they usually smoke a day; then they counted the number of cigarettes smoked the day after hypnosis. (*Note:* It usually takes several weeks for a subject to stop smoking completely, and the method does not work for everyone.) The results follow.

| | Cigarettes Smoked per Day | | | Cigarettes Smoked per Day | |
|---|---|---|---|---|---|
| Subject | After Hypnosis | Before Hypnosis | Subject | After Hypnosis | Before Hypnosis |
| 1 | 28 | 28 | 10 | 5 | 19 |
| 2 | 15 | 35 | 11 | 12 | 32 |
| 3 | 2 | 14 | 12 | 20 | 42 |
| 4 | 20 | 20 | 13 | 30 | 26 |
| 5 | 31 | 25 | 14 | 19 | 37 |
| 6 | 19 | 40 | 15 | 0 | 19 |
| 7 | 6 | 18 | 16 | 16 | 38 |
| 8 | 17 | 15 | 17 | 4 | 23 |
| 9 | 1 | 21 | 18 | 19 | 24 |

Using a 1% level of significance, test the claim that the number of cigarettes smoked per day was less after hypnosis.

10. ***Incomes: Lawyers and Architects*** How do the average weekly incomes of lawyers and architects compare? A random sample of 18 regions in the United States gave the following information about average weekly incomes (in dollars) (Reference: U.S. Department of Labor, Bureau of Labor Statistics).

| Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Lawyers | 709 | 898 | 848 | 1041 | 1326 | 1165 | 1127 | 866 | 1033 |
| Architects | 859 | 936 | 887 | 1100 | 1378 | 1295 | 1039 | 888 | 1012 |

| Region | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|
| Lawyers | 718 | 835 | 1192 | 992 | 1138 | 920 | 1397 | 872 | 1142 |
| Architects | 794 | 900 | 1150 | 1038 | 1197 | 939 | 1124 | 911 | 1171 |

Does this information indicate that architects tend to have a larger average weekly income? Use $\alpha = 0.05$.

11. *High School Dropouts: Male versus Female* Is the high school dropout rate higher for males or females? A random sample of population regions gave the following information about percentage of 15- to 19-year-olds who are high school dropouts (Reference: *Statistical Abstract of the United States,* 121st Edition).

| Region | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Male | 7.3 | 7.5 | 7.7 | 21.8 | 4.2 | 12.2 | 3.5 | 4.2 | 8.0 | 9.7 |
| Female | 7.5 | 6.4 | 6.0 | 20.0 | 2.6 | 5.2 | 3.1 | 4.9 | 12.1 | 10.8 |

| Region | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Male | 14.1 | 3.6 | 3.6 | 4.0 | 5.2 | 6.9 | 15.6 | 6.3 | 8.0 | 6.5 |
| Female | 15.6 | 6.3 | 4.0 | 3.9 | 9.8 | 9.8 | 12.0 | 3.3 | 7.1 | 8.2 |

Does this information indicate that the dropout rates for males and females are different (either way)? Use $\alpha = 0.01$.

12. *Focus Problem: Meteorology* The Focus Problem at the beginning of this chapter asks you to use a sign test with a 5% level of significance to test the claim that the overall temperature distribution of Madison, Wisconsin, is different (either way) from that of Juneau, Alaska. The monthly average data (in °F) are as follows.

| Month | Jan. | Feb. | March | April | May | June |
|---|---|---|---|---|---|---|
| Madison | 17.5 | 21.1 | 31.5 | 46.1 | 57.0 | 67.0 |
| Juneau | 22.2 | 27.3 | 31.9 | 38.4 | 46.4 | 52.8 |

| Month | July | Aug. | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|
| Madison | 71.3 | 69.8 | 60.7 | 51.0 | 35.7 | 22.8 |
| Juneau | 55.5 | 54.1 | 49.0 | 41.5 | 32.0 | 26.9 |

What is your conclusion?

SECTION 11.2

## The Rank-Sum Test

FOCUS POINTS

- State the criteria for setting up a rank-sum test.
- Use the distribution of ranks to complete the test.
- Interpret the results in the context of the application.

The sign test is used when we have paired data values coming from dependent samples, as in "before and after" studies. However, if the data values are *not paired,* the sign test should *not* be used.

Criteria for rank-sum test

For the situation in which we draw *independent random samples* from two populations, there is another nonparametric method for testing the difference between sample means; it is called the *rank-sum test* (also called the *Mann–Whitney test*). The rank-sum test can be used when assumptions about *normal* populations are not satisfied. To fix our thoughts on a definite problem, let's consider the following example:

When a scuba diver makes a deep dive, nitrogen builds up in the diver's blood. After returning to the surface, the diver must wait in a decompression

**TABLE 11-4 Decompression Times for 23 Navy Divers (in min)**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A (had pill) | 41 | 56 | 64 | 42 | 50 | 70 | 44 | 57 | 63 | 65 | 52 | |
| | Mean time = 54.91 min | | | | | | | | | | | |
| Group B (no pill) | 66 | 43 | 72 | 62 | 55 | 80 | 74 | 75 | 77 | 78 | 47 | 60 |
| | Mean time = 65.75 min | | | | | | | | | | | |

chamber until the nitrogen level of the blood returns to normal. A physiologist working with the Navy has invented a pill that a diver takes 1 hour before diving. The pill is supposed to reduce the waiting time spent in the decompression chamber. Twenty-three Navy divers volunteered to help the physiologist determine if the pill has any effect. The divers were randomly divided into two groups: group A had 11 divers who took the pill, and group B had 12 divers who did not take the pill. All the divers worked the same length of time on a deep salvage operation and returned to the decompression chamber. A monitoring device in the decompression chamber measured the waiting time for each diver's nitrogen level to return to normal. These times are recorded in Table 11-4.

Rank the data

The means of our two samples are 54.91 and 65.75 minutes. We will use the rank-sum test to decide whether the difference between the means is significant. First, we arrange the two samples jointly in order of increasing time. To do this, we use the data of groups A and B as if they were one sample. The times (in minutes), groups, and ranks are shown in Table 11-5.

Group A occupies the ranks 1, 2, 4, 6, 7, 9, 10, 13, 14, 15, and 17, while group B occupies the ranks 3, 5, 8, 11, 12, 16, 18, 19, 20, 21, 22, and 23. We add up the ranks of the group with the *smaller* sample size, in this case, group A.

Sum the ranks of the smaller group

The sum of the ranks is denoted by $R$:

$$R = 1 + 2 + 4 + 6 + 7 + 9 + 10 + 13 + 14 + 15 + 17 = 98$$

Let $n_1$ be the size of the *smaller sample* and $n_2$ be the size of the *larger sample*. In the case of the divers, $n_1 = 11$ and $n_2 = 12$. So, $R$ is the sum of the ranks from the smaller sample. If both samples are of the same size, then $n_1 = n_2$ and $R$ is the sum of the ranks of either group (but not both groups).

Andrew G. Wood/Photo Researchers, Inc.

**TABLE 11-5 Ranks for Decompression Time**

| Time | Group | Rank | Time | Group | Rank |
|---|---|---|---|---|---|
| 41 | A | 1 | 63 | A | 13 |
| 42 | A | 2 | 64 | A | 14 |
| 43 | B | 3 | 65 | A | 15 |
| 44 | A | 4 | 66 | B | 16 |
| 47 | B | 5 | 70 | A | 17 |
| 50 | A | 6 | 72 | B | 18 |
| 52 | A | 7 | 74 | B | 19 |
| 55 | B | 8 | 75 | B | 20 |
| 56 | A | 9 | 77 | B | 21 |
| 57 | A | 10 | 78 | B | 22 |
| 60 | B | 11 | 80 | B | 23 |
| 62 | B | 12 | | | |

Distribution of ranks

When both $n_1$ and $n_2$ are sufficiently large (each greater than 10), advanced mathematical statistics can be used to show that $R$ is approximately normally distributed, with mean

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

and standard deviation

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

## GUIDED EXERCISE 3 Mean and standard deviation of ranks

For the Navy divers, compute $\mu_R$ and $\sigma_R$. (Recall that $n_1 = 11$ and $n_2 = 12$.)

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{11(11 + 12 + 1)}{2} = 132$$

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

$$= \sqrt{\frac{11 \cdot 12(11 + 12 + 1)}{12}} \approx 16.25$$

Sample test statistic

Since $n_1 = 11$ and $n_2 = 12$, the samples are large enough to assume that the rank $R$ is approximately normally distributed. We convert the sample test statistic $R$ to a $z$ value using the following formula, with $R = 98$, $\mu_R = 132$, and $\sigma_R \approx 16.25$:

$$z = \frac{R - \mu_R}{\sigma_R} \approx \frac{98 - 132}{16.25} \approx -2.09$$

Hypotheses

When using the rank-sum test, the null hypothesis is that the distributions are the same, while the alternate hypothesis is that the distributions are different. In the case of the Navy divers, we have

$H_0$: Decompression time distributions are the same.

$H_1$: Decompression time distributions are different.

We'll test the decompression time distributions using level of significance 5%.

*P*-value

To find the $P$-value of the sample test statistic $z = -2.09$, we use the normal distribution (Table 5 of Appendix II) and the fact that we have a two-tailed test. Figure 11-3 shows the $P$-value.

The area to the left of $-2.09$ is 0.0183. This is a two-tailed test, so

$$P\text{-value} = 2(0.0183) = 0.0366$$

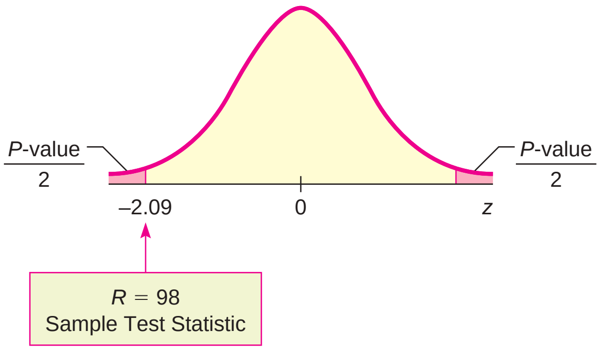
Conclusion

Since the $P$-value is less than $\alpha = 0.05$, we reject $H_0$. At the 5% level of significance, we have sufficient evidence to conclude that the pill changes decompression times for divers.

The steps necessary for a rank-sum test are summarized by the procedure on the next page.

FIGURE 11-3

*P*-value




**PROCEDURE**

## How to construct a rank-sum test

***Setup and Requirements***

You first need independent random samples (both of size 11 or more) from two populations $A$ and $B$. Let $n_1$ be the sample size of the *smaller* sample and let $n_2$ be the sample size of the larger sample. If the sample sizes are equal, then simply use the common value for $n_1$ and $n_2$. Next, you need to rank-order the data as if they were one big sample. Label each rank $A$ or $B$ according to the population from which it came. Let $R$ be a random variable that represents the sum of ranks from the sample of size $n_1$. If $n_1 = n_2$, then $R$ is the sum of ranks from either group (but not both).

***Procedure***

1. Set the *level of significance* $\alpha$. The *null* and *alternate hypotheses* are

   $H_0$: The two samples come from populations with the same distribution (the two populations are identical).

   $H_1$: The two samples come from populations with different distributions (the populations differ in some way).

2. The *sample test statistic* is

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

   where $R =$ sum of ranks from the sample of size $n_1$ (smaller sample),

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

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

   and $n_1 > 10$, $n_2 > 10$.

3. Use the standard normal distribution with a two-tailed test to find the *P-value* corresponding to the test statistic.
4. *Conclude* the test. If $P$-value $\leq \alpha$, then reject $H_0$. If $P$-value $> \alpha$, then do not reject $H_0$.
5. *Interpret your conclusion* in the context of the application.

Procedure for tied ranks

**NOTE** For the decompression time data, there were no ties for any rank. If a tie does occur, then each of the tied observations is given the *mean* of the ranks that it occupies. For example, if we rank the numbers

41 42 44 44 44 44

TABLE 11-6

| Observation | Rank |
|---|---|
| 41 | 1 |
| 42 | 2 |
| 44 | 4.5 |
| 44 | 4.5 |
| 44 | 4.5 |
| 44 | 4.5 |

we see that 44 occupies ranks 3, 4, 5, and 6. Therefore, we give each of the 44's a rank that is the mean of 3, 4, 5, and 6:

$$\text{Mean of ranks} = \frac{3 + 4 + 5 + 6}{4} = 4.5$$

The final ranking would then be that shown in Table 11-6.

For samples where $n_1$ or $n_2$ is less than 11, there are statistical tables that give appropriate critical values for the rank-sum test. Most libraries contain such tables, and the interested reader can find such information by looking under the *Mann–Whitney U Test*.

## GUIDED EXERCISE 4 Rank-sum test

A biologist is doing research on elk in their natural Colorado habitat. Two regions are under study, both having about the same amount of forage and natural cover. However, region A seems to have fewer predators than region B. To determine if there is a difference in elk life spans between the two regions, a sample of 11 mature elk from each region are tranquilized and have a tooth removed. A laboratory examination of the teeth reveals the ages of the elk. Results for each sample are given in Table 11-7. The biologist uses a 5% level of significance to test for a difference in life spans.

TABLE 11-7 **Ages of Elk**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A | 4 | 10 | 11 | 2 | 2 | 3 | 9 | 4 | 12 | 6 | 6 |
| Group B | 7 | 3 | 8 | 4 | 8 | 5 | 6 | 4 | 2 | 4 | 3 |

(a) Fill in the remaining ranks of Table 11-8. Be sure to use the process of taking the mean of tied ranks.

TABLE 11-8 **Ranks of Elk**

| Age | Group | Rank | Age | Group | Rank |
|---|---|---|---|---|---|
| 2 | A | 2 | 5 | B | 12 |
| 2 | A | 2 | 6 | A | — |
| 2 | B | 2 | 6 | A | — |
| 3 | A | 5 | 6 | B | — |
| 3 | B | 5 | 7 | B | — |
| 3 | B | 5 | 8 | B | — |
| 4 | A | 9 | 8 | B | — |
| 4 | A | 9 | 9 | A | — |
| 4 | B | 9 | 10 | A | — |
| 4 | B | 9 | 11 | A | — |
| 4 | B | 9 | 12 | A | — |

| Rank |
|---|
| 12 |
| 14 |
| 14 |
| 14 |
| 16 |
| 17.5 |
| 17.5 |
| 19 |
| 20 |
| 21 |
| 22 |

(b) What is $\alpha$? State the null and alternate hypotheses.

$\alpha = 0.05$

$H_0$: Distributions of life spans are the same.

$H_1$: Distributions of life spans are different.

*Continued*

GUIDED EXERCISE 4 *continued*

(c) Find $\mu_R$, $\sigma_R$, and $R$. Convert $R$ to a sample $z$ statistic.

Since $n_1 = 11$ and $n_2 = 11$,

$$\mu_R = \frac{(11)(11 + 11 + 1)}{2} = 126.5$$

$$\sigma_R = \sqrt{\frac{11 \cdot 11(11 + 11 + 1)}{12}} \approx 15.23$$

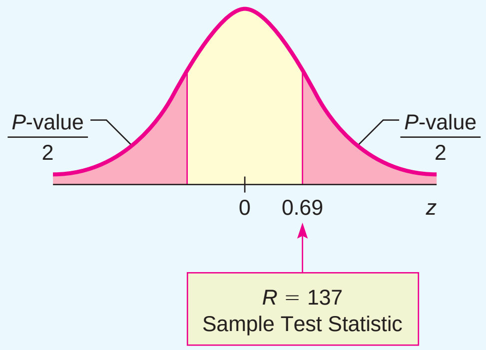
Since $n_1 = n_2 = 11$, we can use the sum of the ranks of either the A group or the B group. Let's use the A group. The A group ranks are 2, 2, 5, 9, 9, 14, 14, 19, 20, 21, and 22. Therefore,

$$R = 2 + 2 + 5 + 9 + 9 + 14 + 14 + 19 + 20 + 21 + 22 = 137$$

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{137 - 126.5}{15.23} \approx 0.69$$

(d) Find the $P$-value shown in Figure 11-4.

FIGURE 11-4 *P*-value




Using Table 5 of Appendix II, the area to the right of 0.69 is 0.2451. Since this is a two-tailed test,

$$P\text{-value} = 2(0.2451) = 0.4902$$

*Comment:* If we use the sum of ranks of group B, then $R_B = 116$ and $z = -0.69$. The $P$-value is again 0.4902, and we have the same conclusion.

(e) ***Interpretation*** What is the conclusion?

The $P$-value of 0.4902 is greater than $\alpha = 0.05$, so we do not reject $H_0$. The evidence does not support the claim that the age distribution of elk is different between the two regions.

**VIEWPOINT** Point Barrow, Alaska

*Point Barrow is located very near the northernmost point of land in the United States. In 1935, Will Rogers (an American humorist, social critic, and philosopher) was killed with Wiley Post (a pioneer aviator) at a landing strip near Point Barrow. Since 1920, a weather station at the (now named) Wiley Post–Will Rogers Memorial Landing Strip has recorded daily high and low temperatures. From these readings, annual mean maximum and minimum temperatures have been computed. Is Point Barrow warming up, cooling down, or neither? Can you think of a way to gather data and construct a nonparametric test to investigate long-term temperature highs and lows at Point Barrow? For weather-related data, visit the Brase/Brase statistics site at* **http://www.cengage.com/statistics/brase** *and find a link to the Geophysical Institute at the University of Alaska in Fairbanks. Then follow the links to Point Barrow.*

## SECTION 11.2 PROBLEMS

1. *Statistical Literacy* When applying the rank-sum test, do you need independent or dependent samples?

2. *Statistical Literacy* If two or more data values are the same, how is the rank of each of the tied data computed?

For Problems 3–11, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Compute the sample test statistic. What is the sampling distribution? What conditions are necessary to use this distribution?
(c) Find the $P$-value of the sample test statistic.
(d) Conclude the test.
(e) ***Interpret*** the conclusion in the context of the application.

3. ***Agriculture: Lima Beans*** Are yields for organic farming different from conventional farming yields? Independent random samples from method A (organic farming) and method B (conventional farming) gave the following information about yield of lima beans (in tons/acre) (Reference: *Agricultural Statistics,* U.S. Department of Agriculture).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method A | 1.83 | 2.34 | 1.61 | 1.99 | 1.78 | 2.01 | 2.12 | 1.15 | 1.41 | 1.95 | 1.25 | |
| Method B | 2.15 | 2.17 | 2.11 | 1.89 | 1.34 | 1.88 | 1.96 | 1.10 | 1.75 | 1.80 | 1.53 | 2.21 |

Use a 5% level of significance to test the hypothesis that there is no difference between the yield distributions.

4. ***Agriculture: Sweet Corn*** Are yields for organic farming different from conventional farming yields? Independent random samples from method A (organic farming) and method B (conventional farming) gave the following information about yield of sweet corn (in tons/acre) (Reference: *Agricultural Statistics,* U.S. Department of Agriculture).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method A | 6.88 | 6.86 | 7.12 | 5.91 | 6.80 | 6.92 | 6.25 | 6.98 | 7.21 | 7.33 | 5.85 | 6.72 |
| Method B | 5.71 | 6.93 | 7.05 | 7.15 | 6.79 | 6.87 | 6.45 | 7.34 | 5.68 | 6.78 | 6.95 | |

Use a 5% level of significance to test the claim that there is no difference between the yield distributions.

5. ***Horse Trainer: Jumps*** A horse trainer teaches horses to jump by using two methods of instruction. Horses being taught by method A have a lead horse that accompanies each jump. Horses being taught by method B have no lead horse. The table shows the number of training sessions required before each horse performed the jumps properly.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method A | 28 | 35 | 19 | 41 | 37 | 31 | 38 | 40 | 25 | 27 | 36 | 43 |
| Method B | 42 | 33 | 26 | 24 | 44 | 46 | 34 | 20 | 48 | 39 | 45 | |

Use a 5% level of significance to test the claim that there is no difference between the training session distributions.

6. ***Violent Crime: FBI Report*** Is the crime rate in New York different from the crime rate in New Jersey? Independent random samples from region A (cities in New York) and region B (cities in New Jersey) gave the following information about violent crime rate (number of violent crimes per 100,000 population) (Reference: U.S. Department of Justice, Federal Bureau of Investigation).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Region A | 554 | 517 | 492 | 561 | 577 | 621 | 512 | 580 | 543 | 605 | 531 | |
| Region B | 475 | 419 | 505 | 575 | 395 | 433 | 521 | 388 | 375 | 411 | 586 | 415 |

Use a 5% level of significance to test the claim that there is no difference in the crime rate distributions of the two states.

7. ***Psychology: Testing*** A cognitive aptitude test consists of putting together a puzzle. Eleven people in group A took the test in a competitive setting (first and second to finish received a prize). Twelve people in group B took the test in a noncompetitive setting. The results follow (in minutes required to complete the puzzle).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A | 7 | 12 | 10 | 15 | 22 | 17 | 18 | 13 | 8 | 16 | 11 | |
| Group B | 9 | 16 | 30 | 11 | 33 | 28 | 19 | 14 | 24 | 27 | 31 | 29 |

Use a 5% level of significance to test the claim that there is no difference in the distributions of time to complete the test.

8. ***Psychology: Testing*** A psychologist has developed a mental alertness test. She wishes to study the effects (if any) of type of food consumed on mental alertness. Twenty-one volunteers were randomly divided into two groups. Both groups were told to eat the amount they usually eat for lunch at noon. At 2:00 P.M., all subjects were given the alertness test. Group A had a low-fat lunch with no red meat, lots of vegetables, carbohydrates, and fiber. Group B had a high-fat lunch with red meat, vegetable oils, and low fiber. The only drink for both groups was water. The test scores are shown below.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A | 76 | 93 | 52 | 81 | 68 | 79 | 88 | 90 | 67 | 85 | 60 | |
| Group B | 44 | 57 | 60 | 91 | 62 | 86 | 82 | 65 | 96 | 42 | 68 | 98 |

Use a 1% level of significance to test the claim that there is no difference in mental alertness distributions based on type of lunch.

9. ***Lifestyles: Exercise*** Is there a link between exercise and level of education? Independent random samples of adults from group A (college graduates) and group B (no high school diploma) gave the following information about percentage who exercise regularly (Reference: Center for Disease Control and Prevention).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A(%) | 63.3 | 55.1 | 50.0 | 47.1 | 58.2 | 60.0 | 44.3 | 49.1 | 68.7 | 57.3 | 59.9 | |
| B(%) | 33.7 | 40.1 | 53.3 | 36.9 | 29.1 | 59.6 | 35.7 | 44.2 | 38.2 | 46.6 | 45.2 | 60.2 |

Use a 1% level of significance to test the claim that there is no difference in the exercise rate distributions according to education level.

10. ***Doctor's Degree: Years of Study*** Does the average length of time to earn a doctorate differ from one field to another? Independent random samples from large graduate schools gave the following averages for length of registered time (in years) from bachelor's degree to doctorate. Sample A was taken from the humanities field, and sample B from the social sciences field (Reference: *Education Statistics*, U.S. Department of Education).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Field A | 8.9 | 8.3 | 7.2 | 6.4 | 8.0 | 7.5 | 7.1 | 6.0 | 9.2 | 8.7 | 7.5 | |
| Field B | 7.6 | 7.9 | 6.2 | 5.8 | 7.8 | 8.3 | 8.5 | 7.0 | 6.3 | 5.4 | 5.9 | 7.7 |

Use a 1% level of significance to test the claim that there is no difference in the distributions of time to complete a doctorate for the two fields.

11. ***Education: Spelling*** Twenty-two fourth-grade children were randomly divided into two groups. Group A was taught spelling by a phonetic method. Group B was taught spelling by a memorization method. At the end of the fourth grade, all children were given a standard spelling exam. The scores are as follows.

| Group A | 77 | 95 | 83 | 69 | 85 | 92 | 61 | 79 | 87 | 93 | 65 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group B | 62 | 90 | 70 | 81 | 63 | 75 | 80 | 72 | 82 | 94 | 65 | 79 |

Use a 1% level of significance to test the claim that there is no difference in the test score distributions based on instruction method.

SECTION 11.3

## Spearman Rank Correlation

FOCUS POINTS

- Learn about monotone relations and the Spearman rank correlation coefficient.
- Compute the Spearman correlation coefficient and conduct statistical tests for significance.
- Interpret the results in the context of the application.

Data given in ranked form (ordinal type) are different from data given in measurement form (interval or ratio type). For instance, if we compared the test performances of three students and, say, Elizabeth did the best, Joel did next best, and Sally did the worst, we are giving the information in ranked form. We cannot say how much better Elizabeth did than Sally or Joel, but we do know how the three scores compare. If the actual test scores for the three tests were given, we would have data in measurement form and could tell exactly how much better Elizabeth did than Joel or Sally. In Chapter 9, we studied linear correlation of data in measurement form. In this section, we will study correlation of data in ranked form.

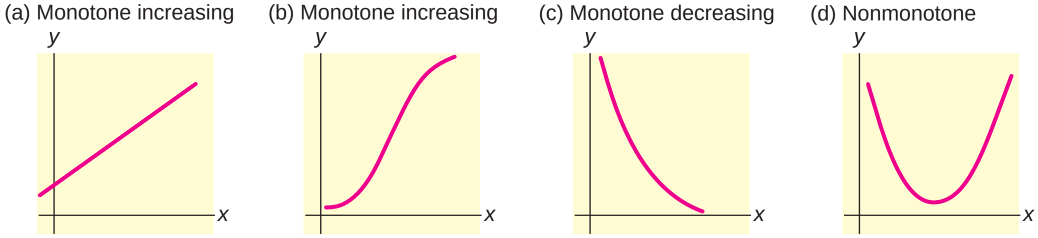
As a specific example of a situation in which we might want to compare ranked data from two sources, consider the following. Hendricks College has a new faculty position in its political science department. A national search to fill this position has resulted in a large number of qualified candidates. The political science faculty reserves the right to make the final hiring decision. However, the faculty is interested in comparing its opinion with student opinion about the teaching ability of the candidates. A random sample of nine equally qualified candidates were asked to give a classroom presentation to a large class of students. Both faculty and students attended the lectures. At the end of each lecture, both faculty and students filled out a questionnaire about the teaching performance of the candidate. Based on these questionnaires, each candidate was given an overall rank from the faculty and an overall rank from the students. The results are shown in Table 11-9. Higher ranks mean better teaching performance.

**TABLE 11-9** **Faculty and Student Ranks of Candidates**

| Candidate | Faculty Rank | Student Rank |
|---|---|---|
| 1 | 3 | 5 |
| 2 | 7 | 7 |
| 3 | 5 | 6 |
| 4 | 9 | 8 |
| 5 | 2 | 3 |
| 6 | 8 | 9 |
| 7 | 1 | 1 |
| 8 | 6 | 4 |
| 9 | 4 | 2 |

FIGURE 11-5

Examples of Monotone Relations




Using data in ranked form, we answer the following questions:

1. Do candidates getting higher ranks from faculty tend to get higher ranks from students?
2. Is there any relation between faculty rankings and student rankings?
3. Do candidates getting higher ranks from faculty tend to get lower ranks from students?

Monotone relationship

We will use the Spearman rank correlation to answer such questions. In the early 1900s, Charles Spearman of the University of London developed the techniques that now bear his name. The Spearman test of rank correlation requires us to use *ranked variables*. Because we are using only ranks, we cannot use the Spearman test to check for the existence of a linear relationship between the variables as we did with the Pearson correlation coefficient (Section 9.1). The Spearman test checks only for the existence of a *monotone* relationship between the variables. (See Figure 11-5.) By a *monotone relationship** between variables $x$ and $y$, we mean a relationship in which

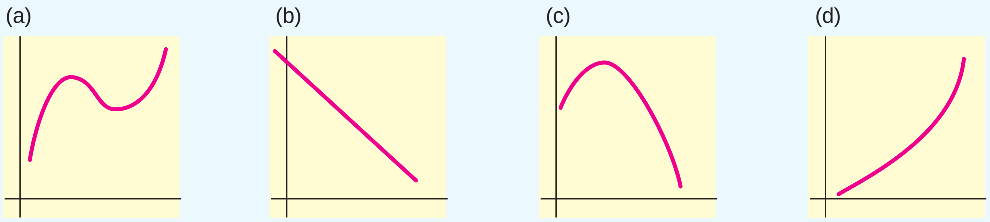
1. as $x$ increases, $y$ also increases, or
2. as $x$ increases, $y$ decreases.

The relationship shown in Figure 11-5(d) is a nonmonotone relationship because as $x$ increases, $y$ at first decreases, but later starts to increase. Remember, for a relation to be monotone, as $x$ increases, $y$ must *always* increase or *always* decrease. In a nonmonotone relationship, as $x$ increases, $y$ sometimes increases and sometimes decreases or stays unchanged.

**GUIDED EXERCISE 5** **Monotonic behavior**

Identify each of the relations in Figure 11-6 as monotone increasing, monotone decreasing, or nonmonotone.

FIGURE 11-6




*Answers:* (a) nonmonotone, (b) monotone decreasing, (c) nonmonotone, (d) monotone increasing

*Some advanced texts call the monotone relationship we describe *strictly monotone*.

Before we can complete the solution of our problem about the political science department at Hendricks College, we need the following information.

Suppose we have a sample of size $n$ of randomly obtained ordered pairs $(x, y)$, where both the $x$ and $y$ values are from *ranked variables*. If there are no ties in the ranks, then the Pearson product-moment correlation coefficient (Section 9.1) can be reduced to a simpler equation. The new equation produces the *Spearman rank correlation coefficient*, $r_s$.

Spearman rank correlation coefficient $r_s$

**Spearman rank correlation coefficient**

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} \quad \text{where } d = x - y$$

The Spearman rank correlation coefficient has the following properties.

**Properties of the Spearman rank correlation coefficient**

1. $-1 \le r_s \le 1$. If $r_s = -1$, the relation between $x$ and $y$ is perfectly mono tone decreasing. If $r_s = 0$, there is no monotone relation between $x$ and $y$. If $r_s = 1$, the relation between $x$ and $y$ is perfectly monotone increasing. Values of $r_s$ close to 1 or $-1$ indicate a strong tendency for $x$ and $y$ to have a monotone relationship (increasing or decreasing). Values of $r_s$ close to 0 indicate a very weak (or perhaps nonexistent) monotone relationship.
2. The probability distribution of $r_s$ depends on the sample size $n$. It is symmetric about $r_s = 0$. Table 9 of Appendix II gives critical values for certain specified one-tail and two-tail areas. Use of the table requires no assumptions that $x$ and $y$ are normally distributed variables. In addition, we make no assumption about the $x$ and $y$ relationship being linear.
3. The Spearman rank correlation coefficient $r_s$ is the *sample* estimate for the *population* Spearman rank correlation coefficient $\rho_s$.

Population Spearman rank correlation coefficient $\rho_s$

We construct a test of significance for the Spearman rank correlation coefficient in much the same way that we tested the Pearson correlation coefficient (Section 9.3). The null hypothesis states that there is no monotone relation between $x$ and $y$ (either increasing or decreasing).

Hypotheses

$H_0$: $\rho_s = 0$

The alternate hypothesis is one of the following:

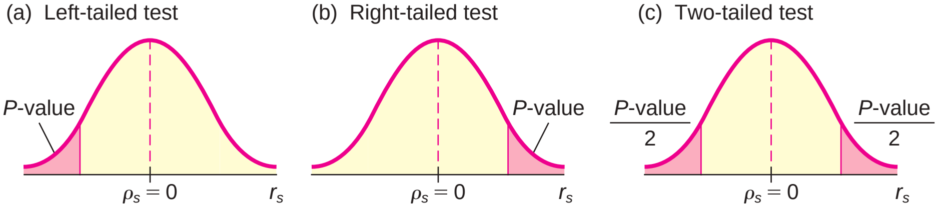
| | | |
|---|---|---|
| $H_1$: $\rho_s < 0$ | $H_1$: $\rho_s > 0$ | $H_1$: $\rho_s \ne 0$ |
| (left-tailed) | (right-tailed) | (two-tailed) |

A left-tailed alternate hypothesis claims there is a monotone-decreasing relation between $x$ and $y$. A right-tailed alternate hypothesis claims there is a monotone-increasing relation between $x$ and $y$, while a two-tailed alternate hypothesis claims there is a monotone relation (either increasing or decreasing) between $x$ and $y$.

Figure 11-7 shows the type of test and corresponding $P$-value region.

**FIGURE 11-7**

**Type of Test and *P*-value Region**




## EXAMPLE 1 Testing the Spearman rank correlation coefficient

Using the information about the Spearman rank correlation coefficient, let's finish our problem about the search for a new member of the political science department at Hendricks College. Our work is organized in Table 11-10, where the rankings given by students and faculty are listed for each of the nine candidates.

Zigy Kaluzny-Charles Thatcher/Stone/Getty Images

*P*-value

(a) Using a 1% level of significance, let's test the claim that the faculty and students tend to agree about a candidate's teaching ability. This means that the $x$ and $y$ variables should be monotone increasing (as $x$ increases, $y$ increases). Since $\rho_s$ is the population Spearman rank correlation coefficient, we have

$H_0$: $\rho_s = 0$ (There is no monotone relation.)

$H_1$:$\rho_s > 0$ (There is a monotone-increasing relation.)

(b) Compute the sample test statistic.

**SOLUTION:** Since the sample size is $n = 9$, and from Table 11-10 we see that $\Sigma d^2 = 16$, the Spearman rank correlation coefficient is

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(16)}{9(81 - 1)} \approx 0.867$$

(c) Find or estimate the *P*-value.

**SOLUTION:** To estimate the *P*-value for the sample test statistic $r_s = 0.867$, we use Table 9 of Appendix II. The sample size is $n = 9$ and the test is a one-tailed test. We find the location of the sample test statistic in row 9, and then read the corresponding one-tail area. From the Table 9, Appendix II excerpt, we see that the sample test statistic $r_s = 0.867$ falls between the entries 0.834 and 0.917 in the $n = 9$ row. These values correspond to *one-tail areas* between 0.005 and 0.001.

$0.001 < P\text{-value} < 0.005$

**TABLE 11-10 Student and Faculty Ranks of Candidates and Calculations for the Spearman Rank Correlation Test**

| Candidate | Faculty Rank $x$ | Student Rank $y$ | $d = x - y$ | $d^2$ |
|---|---|---|---|---|
| 1 | 3 | 5 | −2 | 4 |
| 2 | 7 | 7 | 0 | 0 |
| 3 | 5 | 6 | −1 | 1 |
| 4 | 9 | 8 | 1 | 1 |
| 5 | 2 | 3 | −1 | 1 |
| 6 | 8 | 9 | −1 | 1 |
| 7 | 1 | 1 | 0 | 0 |
| 8 | 6 | 4 | 2 | 4 |
| 9 | 4 | 2 | 2 | 4 |
| | | | | $\Sigma d^2 = 16$ |

(d) Conclude the test. ***Interpret*** the results.

**SOLUTION:**


| ✓ One-tail area | 0.005 | 0.001 |
|---|---|---|
| $n = 9$ | 0.834 | 0.917 |
| | ↑ Sample $r_s = 0.867$ | |

Since the $P$-value is less than $\alpha = 0.01$, we reject $H_0$. At the 1% level of significance, we conclude that the relation between faculty and student ratings is monotone increasing. This means that faculty and students tend to rank the teaching performance of candidates in a similar way: Higher student ratings of a candidate correspond with higher faculty ratings of the same candidate.

The following procedure summarizes the steps involved in testing the population Spearman rank correlation coefficient.

**PROCEDURE**

## HOW TO TEST THE SPEARMAN RANK CORRELATION COEFFICIENT $\rho_s$

***Setup***

You first need a random sample (of size $n$) of data pairs $(x, y)$, where both the $x$ and $y$ values are *ranked* variables. Let $\rho_s$ represent the population Spearman rank correlation coefficient, which is in theory computed from the population of all possible $(x, y)$ data pairs.

***Procedure***

1. Set the *level of significance* $\alpha$. The *null hypothesis* is $H_0$: $\rho_s = 0$. In the context of the application, choose the *alternate hypothesis* to be $H_1$: $\rho_s > 0$ or $H_1$: $\rho_s < 0$ or $H_1$: $\rho_s \neq 0$.
2. If there are no ties in the ranks, or if the number of ties is small compared to the number of data pairs $n$, then compute the *sample test statistic*

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

   where $d = x - y$ is the difference in ranks

   $n =$ number of data pairs

   and the sum is over all sample data pairs.
3. Use Table 9 of Appendix II to find or estimate the *P-value* corresponding to $r_s$ and $n =$ number of data pairs.
4. *Conclude* the test. If $P$-value $\leq \alpha$, then reject $H_0$. If $P$-value $> \alpha$, then do not reject $H_0$.
5. *Interpret your conclusion* in the context of the application.

**GUIDED EXERCISE 6** | **Testing the Spearman rank correlation coefficient**

Fishermen in the Adirondack Mountains are complaining that acid rain caused by air pollution is killing fish in their region. To research this claim, a team of biologists studied a random sample of 12 lakes in the region. For each lake, they measured the level of acidity of rain in the drainage leading into the lake and the density of fish in the lake (number of fish per acre-foot of water). They then did a ranking of $x =$ acidity and $y =$ density of fish. The results are shown in Table 11-11. Higher $x$ ranks mean more acidity, and higher $y$ ranks mean higher density of fish.

*Continued*

GUIDED EXERCISE 6 *continued*

**Table 11-11 Acid Rain and Density of Fish**

| Lake | Acidity $x$ | Fish Density $y$ | $d = x - y$ | $d^2$ |
|---|---|---|---|---|
| 1 | 5 | 8 | −3 | 9 |
| 2 | 8 | 6 | 2 | 4 |
| 3 | 3 | 9 | −6 | 36 |
| 4 | 2 | 12 | −10 | 100 |
| 5 | 6 | 7 | −1 | 1 |
| 6 | 1 | 10 | −9 | 81 |
| 7 | 10 | 2 | 8 | 64 |
| 8 | 12 | 1 | ___ | ___ |
| 9 | 7 | 5 | ___ | ___ |
| 10 | 4 | 11 | ___ | ___ |
| 11 | 9 | 4 | ___ | ___ |
| 12 | 11 | 3 | ___ | ___ |
| | | | $\Sigma d^2 =$ | ___ |

(a) Complete the entries in the $d$ and $d^2$ columns of Table 11-11, and find $\Sigma d^2$.

| Lake | $x$ | $y$ | $d$ | $d^2$ |
|---|---|---|---|---|
| 8 | 12 | 1 | 11 | 121 |
| 9 | 7 | 5 | 2 | 4 |
| 10 | 4 | 11 | −7 | 49 |
| 11 | 9 | 4 | 5 | 25 |
| 12 | 11 | 3 | 8 | 64 |
| | | | | $\Sigma d^2 = 558$ |

(b) Compute $r_s$.

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(558)}{12(144 - 1)} \approx -0.951$$

(c) The fishermen claim that more acidity means lower density of fish. Does this claim state that $x$ and $y$ have a monotone-increasing relation, a monotone-decreasing relation, or no monotone relation?

The claim states that as $x$ increases, $y$ decreases, so the relation between $x$ and $y$ is monotone decreasing.

(d) To test the fishermen's claim, what should we use for the null hypothesis and for the alternate hypothesis? Use $\alpha = 0.01$.

$H_0$: $\rho_s = 0$ (no monotone relation)
$H_1$: $\rho_s < 0$ monotone-decreasing relation)

(e) Find or estimate the $P$-value of the sample test statistic $r_s = -0.951$.

Use Table 9 of Appendix II. There are $n = 12$ data pairs. The sample statistic $r_s$ is negative. Because the $r_s$ distribution is symmetric about 0, we look up the corresponding positive value 0.951 in the row headed by $n = 12$. Use one-tail areas, since this is a left-tailed test.

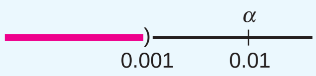
| ✓ One-tail area | 0.001 |
|---|---|
| $n = 12$ | 0.826 |
| | ↑ $-r_s = 0.951$ |

As positive $r_s$ values increase, corresponding right-tail areas decrease. Therefore,

$P$-value $< 0.001$

GUIDED EXERCISE 6 *continued*

(f) Use $\alpha = 0.01$ and conclude the test.




Since the $P$-value is less than $\alpha = 0.01$, we reject $H_0$ and conclude that there is a monotone-decreasing relationship between the acidity of the water and the number of fish.

(g) ***Interpretation*** Do the data support the claim that higher acidity means fewer fish?

At the 1% level of significance, we conclude that higher acidity means fewer fish.

Ties of ranks

If ties occur in the assignment of ranks, we follow the usual method of averaging tied ranks. This method was discussed in Section 11.2 (The Rank-Sum Test). The next example illustrates the method.

COMMENT Technically, the use of the given formula for $r_s$ requires that there be no ties in rank. However, if the number of ties in rank is small relative to the number of ranks, the formula can be used with quite a bit of reliability.

## EXAMPLE 2 TIED RANKS

Do people who smoke more tend to drink more cups of coffee? The following data were obtained from a random sample of $n = 10$ cigarette smokers who also drink coffee.

| Person | Cigarettes per Day | Cups of Coffee per Day |
|---|---|---|
| 1 | 8 | 4 |
| 2 | 15 | 7 |
| 3 | 20 | 10 |
| 4 | 5 | 3 |
| 5 | 22 | 9 |
| 6 | 15 | 5 |
| 7 | 15 | 8 |
| 8 | 25 | 11 |
| 9 | 30 | 18 |
| 10 | 35 | 18 |

(a) To use the Spearman rank correlation test, we need to rank the data. It does not matter if we rank from smallest to largest or from largest to smallest. The only requirement is that we be consistent in our rankings. Let us rank from smallest to largest.

First, we rank the data for each variable as though there were no ties; then we average the ties as shown in Tables 11-12 and 11-13.

(b) Using 0.01 as the level of significance, we test the claim that $x$ and $y$ have a monotone-increasing relationship. In other words, we test the claim that people who tend to smoke more tend to drink more cups of coffee (Table 11-14).

$H_0$: $\rho_s = 0$ (There is no montone relation.)

$H_1$: $\rho_s > 0$ (Right-tailed test)

**TABLE 11-12 Rankings of Cigarettes Smoked per Day**

| Person | Cigarettes per Day | | Rank | | Average Rank $x$ | |
|---|---|---|---|---|---|---|
| 4 | 5 | | 1 | | 1 | |
| 1 | 8 | | 2 | | 2 | |
| 2 | 15 | Ties | 3 | Average rank is 4. | 4 | Use the average rank for tied data. |
| 6 | 15 | | 4 | | 4 | |
| 7 | 15 | | 5 | | 4 | |
| 3 | 20 | | 6 | | 6 | |
| 5 | 22 | | 7 | | 7 | |
| 8 | 25 | | 8 | | 8 | |
| 9 | 30 | | 9 | | 9 | |
| 10 | 35 | | 10 | | 10 | |

**TABLE 11-13 Rankings of Cups of Coffee per Day**

| Person | Cups of Coffee per Day | | Rank | | Average Rank $y$ | |
|---|---|---|---|---|---|---|
| 4 | 3 | | 1 | | 1 | |
| 1 | 4 | | 2 | | 2 | |
| 6 | 5 | | 3 | | 3 | |
| 2 | 7 | | 4 | | 4 | |
| 7 | 8 | | 5 | | 5 | |
| 5 | 9 | | 6 | | 6 | |
| 3 | 10 | | 7 | | 7 | |
| 8 | 11 | | 8 | | 8 | |
| 9 | 18 | Ties | 9 | Average rank is 9.5. | 9.5 | Use the average rank for tied data. |
| 10 | 18 | | 10 | | 9.5 | |

**TABLE 11-14 Ranks to Be Used for a Spearman Rank Correlation Test**

| Person | Cigarette Rank $x$ | Coffee Rank $y$ | $d = x - y$ | $d^2$ |
|---|---|---|---|---|
| 1 | 2 | 2 | 0 | 0 |
| 2 | 4 | 4 | 0 | 0 |
| 3 | 6 | 7 | −1 | 1 |
| 4 | 1 | 1 | 0 | 0 |
| 5 | 7 | 6 | 1 | 1 |
| 6 | 4 | 3 | 1 | 1 |
| 7 | 4 | 5 | −1 | 1 |
| 8 | 8 | 8 | 0 | 0 |
| 9 | 9 | 9.5 | −0.5 | 0.25 |
| 10 | 10 | 9.5 | 0.5 | 0.25 |
| | | | | $\Sigma d^2 = 4.5$ |

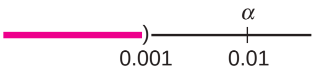
(c) Next, we compute the observed sample test statistic $r_s$ using the results shown in Table 11-14.

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(4.5)}{10(100 - 1)} \approx 0.973$$

(d) Find or estimate the $P$-value for the sample test statistic $r_s = 0.973$.
We use Table 9 of Appendix II to estimate the $P$-value. Using $n = 10$ and a one-tailed test, we see that $r_s = 0.973$ is to the right of the entry 0.879. Therefore, the $P$-value is smaller than 0.001.

| ✓ One-tail area | 0.001 |
|---|---|
| $n = 9$ | 0.879 |
| | ↑ Sample $r_s = 0.973$ |

(e) Conclude the test and *interpret* the results.




Since the $P$-value is less than $\alpha = 0.01$, we reject $H_0$. At the 1% level of significance, it appears that there is a monotone-increasing relationship between the number of cigarettes smoked and the amount of coffee consumed. People who smoke more cigarettes tend to drink more coffee.

**VIEWPOINT** Rug Rats!

*When do babies start to crawl? Janette Benson, in her article "Infant Behavior and Development," claims that crawling age is related to temperature during the month in which babies first try to crawl. To find a data file for this subject, visit the Brase/Brase statistics site at* **http://www.cengage.com/statistics/brase** *and find the link to DASL, the Carnegie Mellon University Data and Story Library. Then look under Psychology in the Data Subjects and select the Crawling Datafile. Can you think of a way to gather data and construct a nonparametric test to study this claim?*

## SECTION 11.3 PROBLEMS

1. *Statistical Literacy* For data pairs $(x, y)$, if $y$ always increases as $x$ increases, is the relationship monotone increasing, monotone decreasing, or nonmonotone?

2. *Statistical Literacy* Consider the Spearman rank correlation coefficient $r_s$ for data pairs $(x, y)$. What is the monotone relationship, if any, between $x$ and $y$ implied by a value of
(a) $r_s = 0$?
(b) $r_s$ close to 1?
(c) $r_s$ close to $-1$?

For Problems 3–11, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Compute the sample test statistic.
(c) Find or estimate the $P$-value of the sample test statistic.
(d) Conclude the test.
(e) *Interpret* the conclusion in the context of the application.

3. ***Training Program: Sales*** A data-processing company has a training program for new salespeople. After completing the training program, each trainee is ranked by his or her instructor. After a year of sales, the same class of trainees is again ranked by a company supervisor according to net value of the contracts they have acquired for the company. The results for a random sample of 11 salespeople trained in the previous year follow, where $x$ is rank in training class and $y$ is rank in sales after 1 year. Lower ranks mean higher standing in class and higher net sales.

| Person | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ rank | 6 | 8 | 11 | 2 | 5 | 7 | 3 | 9 | 1 | 10 | 4 |
| $y$ rank | 4 | 9 | 10 | 1 | 6 | 7 | 8 | 11 | 3 | 5 | 2 |

Using a 0.05 level of significance, test the claim that the relation between $x$ and $y$ is monotone (either increasing or decreasing).

4. ***Economics: Stocks*** As an economics class project, Debbie studied a random sample of 14 stocks. For each of these stocks, she found the cost per share (in dollars) and ranked each of the stocks according to cost. After 3 months, she found the earnings per share for each stock (in dollars). Again, Debbie ranked each of the stocks according to earnings. The way Debbie ranked, higher ranks mean higher cost and higher earnings. The results follow, where $x$ is the rank in cost and $y$ is the rank in earnings.

| Stock | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ rank | 5 | 2 | 4 | 7 | 11 | 8 | 12 | 3 | 13 | 14 | 10 | 1 | 9 | 6 |
| $y$ rank | 5 | 13 | 1 | 10 | 7 | 3 | 14 | 6 | 4 | 12 | 8 | 2 | 11 | 9 |

Using a 0.01 level of significance, test the claim that there is a monotone relation, either way, between the ranks of cost and earnings.

5. ***Psychology: Rat Colonies*** A psychology professor is studying the relation between overcrowding and violent behavior in a rat colony. Eight colonies with different degrees of overcrowding are being studied. By using a television monitor, lab assistants record incidents of violence. Each colony has been ranked for crowding and violence. A rank of 1 means most crowded or most violent. The results for the eight colonies are given in the following table, with $x$ being the population density rank and $y$ the violence rank.

| Colony | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x$ rank | 3 | 5 | 6 | 1 | 8 | 7 | 4 | 2 |
| $y$ rank | 1 | 3 | 5 | 2 | 8 | 6 | 4 | 7 |

Using a 0.05 level of significance, test the claim that lower crowding ranks mean lower violence ranks (i.e., the variables have a monotone-increasing relationship).

6. ***FBI Report: Murder and Arson*** Is there a relation between murder and arson? A random sample of 15 Midwest cities (over 10,000 population) gave the following information about annual number of murder and arson cases (Reference: Federal Bureau of Investigation, U.S. Department of Justice).

| City | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Murder | 12 | 7 | 25 | 4 | 10 | 15 | 9 | 8 | 11 | 18 | 23 | 19 | 21 | 17 | 6 |
| Arson | 62 | 12 | 153 | 2 | 63 | 93 | 31 | 29 | 47 | 131 | 175 | 129 | 162 | 115 | 4 |

(i) Rank-order murder using 1 as the largest data value. Also rank-order arson using 1 as the largest data value. Then construct a table of ranks to be used for a Spearman rank correlation test.
(ii) Use a 1% level of significance to test the claim that there is a monotone-increasing relationship between the ranks of murder and arson.

7. *Psychology: Testing* An army psychologist gave a random sample of seven soldiers a test to measure sense of humor and another test to measure aggressiveness. Higher scores mean greater sense of humor or more aggressiveness.

| Soldier | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Score on humor test | 60 | 85 | 78 | 90 | 93 | 45 | 51 |
| Score on aggressiveness test | 78 | 42 | 68 | 53 | 62 | 50 | 76 |

(i) Ranking the data with rank 1 for highest score on a test, make a table of ranks to be used in a Spearman rank correlation test.
(ii) Using a 0.05 level of significance, test the claim that rank in humor has a monotone-decreasing relation to rank in aggressiveness.

8. *FBI Report: Child Abuse and Runaway Children* Is there a relation between incidents of child abuse and number of runaway children? A random sample of 15 cities (over 10,000 population) gave the following information about the number of reported incidents of child abuse and the number of runaway children (Reference: Federal Bureau of Investigation, U.S. Department of Justice).

| City | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abuse cases | 49 | 74 | 87 | 10 | 26 | 119 | 35 | 13 | 89 | 45 | 53 | 22 | 65 | 38 | 29 |
| Runaways | 382 | 510 | 581 | 163 | 210 | 791 | 275 | 153 | 491 | 351 | 402 | 209 | 410 | 312 | 210 |

(i) Rank-order abuse using 1 as the largest data value. Also rank-order runaways using 1 as the largest data value. Then construct a table of ranks to be used for a Spearman rank correlation test.
(ii) Use a 1% level of significance to test the claim that there is a monotone-increasing relationship between the ranks of incidents of abuse and number of runaway children.

9. *Demographics: Police and Fire Protection* Is there a relation between police protection and fire protection? A random sample of large population areas gave the following information about the number of local police and the number of local firefighters (units in thousands) (Reference: *Statistical Abstract of the United States*).

| Area | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Police | 11.1 | 6.6 | 8.5 | 4.2 | 3.5 | 2.8 | 5.9 | 7.9 | 2.9 | 18.0 | 9.7 | 7.4 | 1.8 |
| Firefighters | 5.5 | 2.4 | 4.5 | 1.6 | 1.7 | 1.0 | 1.7 | 5.1 | 1.3 | 12.6 | 2.1 | 3.1 | 0.6 |

(i) Rank-order police using 1 as the largest data value. Also rank-order firefighters using 1 as the largest data value. Then construct a table of ranks to be used for a Spearman rank correlation test.
(ii) Use a 5% level of significance to test the claim that there is a monotone relationship (either way) between the ranks of number of police and number of firefighters.

10. *Ecology: Wetlands* Turbid water is muddy or cloudy water. Sunlight is necessary for most life forms; thus turbid water is considered a threat to wetland ecosystems. Passive filtration systems are commonly used to reduce turbidity in wetlands. Suspended solids are measured in mg/l. Is there a relation between input and output turbidity for a passive filtration system and, if so, is

it statistically significant? At a wetlands environment in Illinois, the inlet and outlet turbidity of a passive filtration system have been measured. A random sample of measurements is shown below (Reference: *EPA Wetland Case Studies*).

| Reading | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Inlet (mg/l) | 8.0 | 7.1 | 24.2 | 47.7 | 50.1 | 63.9 | 66.0 | 15.1 | 37.2 | 93.1 | 53.7 | 73.3 |
| Outlet (mg/l) | 2.4 | 3.6 | 4.5 | 14.9 | 7.4 | 7.4 | 6.7 | 3.6 | 5.9 | 8.2 | 6.2 | 18.1 |

(i) Rank-order the inlet readings using 1 as the largest data value. Also rank-order the outlet readings using 1 as the largest data value. Then construct a table of ranks to be used for a Spearman rank correlation test.

(ii) Use a 1% level of significance to test the claim that there is a monotone relation ship (either way) between the ranks of the inlet readings and outlet readings.

11. *Insurance: Sales* Big Rock Insurance Company did a study of per capita income and volume of insurance sales in eight Midwest cities. The volume of sales in each city was ranked, with 1 being the largest volume. The per capita income was rounded to the nearest thousand dollars.

| City | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Rank of insurance sales volume | 6 | 7 | 1 | 8 | 3 | 2 | 5 | 4 |
| Per capita income in \$1000 | 17 | 18 | 19 | 11 | 16 | 20 | 15 | 19 |

(i) Using a rank of 1 for the highest per capita income, make a table of ranks to be used for a Spearman rank correlation test.

(ii) Using a 0.01 level of significance, test the claim that there is a monotone relation (either way) between rank of sales volume and rank of per capita income.

## SECTION 11.4 Runs Test for Randomness

**FOCUS POINTS**

- Test a sequence of *symbols* for randomness.
- Test a sequence of *numbers* for randomness about the median.

Astronomers have made an extensive study of galaxies that are ±16° above and below the celestial equator. Of special interest is the flux, or change in radio signals, that originates from large electromagnetic disturbances deep in space. The flux units ($10^{-26}$ watts/m$^2$/Hz) are very small. However, modern radio astronomy can detect and analyze these signals using large antennas (Reference: *Journal of Astrophysics,* Vol. 148, pp. 321–365).

A very important question is the following: Are changes in flux simply random, or is there some kind of nonrandom pattern? Let us use the symbol S to represent a strong or moderate flux and the symbol W to represent a faint or weak flux. Astronomers have received the following signals in order of occurrence.

S S W W W S W W S S S W W W S S W W W S S

Is there a statistical test to help us decide whether or not this sequence of radio signals is random? Well, we're glad you asked, because that is the topic of this section.

We consider applications in which *two* symbols are used (e.g., S or W). Applications using more than two symbols are left to specialized studies in mathematical combinatorics.

Sequence

Run

A **sequence** is an *ordered set* of consecutive symbols.

A **run** is a sequence of one or more occurrences of the *same* symbol.

$n_1$ = number of times the first symbol occurs in a sequence

$n_2$ = number of times the second symbol occurs in a sequence

$R$ is a random variable that represents the **number of runs in a sequence.**

### EXAMPLE 3 BASIC TERMINOLOGY

In this example, we use the symbols S and W, where S is the first symbol and W is the second symbol, to demonstrate sequences and runs. Identify the runs.

(a) S S W W W is a sequence.

**SOLUTION:** Table 11-15 shows the sequence of runs. There are $R = 2$ runs in the sequence. The first symbol S occurs $n_1 = 2$ times. The second symbol W occurs $n_2 = 3$ times.

**TABLE 11-15 Runs**

| Run 1 | Run 2 |
|---|---|
| S S | W W W |

(b) S S W W W S W W S S S S W is a sequence.

**SOLUTION:** The sequence of runs are shown in Table 11-16. There are $R = 6$ runs in the sequence. The first symbol S occurs $n_1 = 7$ times. The second symbol W occurs $n_2 = 6$ times.

**TABLE 11-16 Runs**

| Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Run 6 |
|---|---|---|---|---|---|
| S S | W W W | S | W W | S S S S | W |

Hypotheses

To test a sequence of two symbols for randomness, we use the following hypotheses.

Hypotheses for runs test for randomness

Hypotheses for runs test for randomness

$H_0$: The symbols are randomly mixed in the sequence.

$H_1$: The symbols are not randomly mixed in the sequence.

The decision procedure will reject $H_0$ if either $R$ is too small (too few runs) or $R$ is too large (too many runs).

Sample test statistic

The number of runs $R$ is a *sample test statistic* with its own sampling distribution. Table 10 of Appendix II gives critical values of $R$ for a significance level $\alpha = 0.05$. There are two parameters associated with $R$. They are $n_1$ and $n_2$, the numbers of times the first and second symbols appear in the sequence, respectively. If either $n_1 > 20$ or $n_2 > 20$, you can apply the normal approximation to construct the test. This will be discussed in Problems 11 and 12 at the end of this section. For now, we assume that $n_1 \leq 20$ and $n_2 \leq 20$.

Critical values

For each pair of $n_1$ and $n_2$ values, Table 10 of Appendix II provides two critical values: a smaller value denoted $c_1$ and a larger value denoted $c_2$. These two values are used to decide whether or not to reject the null hypothesis $H_0$ that the symbols are randomly mixed in the sequence.

**Decision process when $n_1 \le 20$ and $n_2 \le 20$**

Use Table 10 of Appendix II with $n_1$ and $n_2$ to find the critical values $c_1$ and $c_2$. At the $\alpha = 5\%$ level of significance, use the following decision process, where $R$ is the number of runs: If either $R \le c_1$ (too few runs) or $R \ge c_2$ (too many runs), then *reject* $H_0$. Otherwise, *do not reject* $H_0$.

COMMENT If either $n_1$ or $n_2$ is larger than 20, a normal approximation can be used. See Problems 11 and 12 at the end of this section.

Let's apply this decision process to the astronomy example regarding the sequence of strong and weak electromagnetic radio signals coming from a distant galaxy.

**EXAMPLE 4** RUNS TEST

Recall that our astronomers had received the following sequence of electromagnetic signals, where S represents a strong flux and W represents a weak flux.

S S W W W S W W S S S W W W S S W W W S S

Is this a random sequence or not? Use a 5% level of significance.

(a) What is the level of significance $\alpha$? State the null and alternate hypotheses.

**SOLUTION:** $\alpha = 0.05$

$H_0$: The symbols S and W are randomly mixed in the sequence.

$H_1$: The symbols S and W are not randomly mixed in the sequence.

(b) Find the sample test statistic $R$ and the parameters $n_1$ and $n_2$.

**SOLUTION:** We break the sequence according to runs.

| Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Run 6 | Run 7 | Run 8 | Run 9 |
|---|---|---|---|---|---|---|---|---|
| SS | WWW | S | WW | SSS | WWW | SS | WWW | SS |

We see that there are $n_1 = 10$ S symbols and $n_2 = 11$ W symbols. The number of runs is $R = 9$.

(c) Use Table 10 of Appendix II to find the critical values $c_1$ and $c_2$.

**SOLUTION:** Since $n_1 = 10$ and $n_2 = 11$, then $c_1 = 6$ and $c_2 = 17$.

Eastcott-Momatiuk/The Image Works

(d) Conclude the test.

**SOLUTION:**

| $R \le 6$ | ✓ $7 \le R \le 16$ | $R \ge 17$ |
|---|---|---|
| Reject $H_0$. | ✓ Fail to reject $H_0$. | Reject $H_0$. |

Since $R = 9$, we fail to reject $H_0$ at the 5% level of significance.

(e) *Interpret* the conclusion in the context of the problem.

**SOLUTION:** At the 5% level of significance, there is insufficient evidence to conclude that the sequence of electromagnetic signals is not random.

Randomness about the median

An important application of the runs test is to help us decide if a sequence of numbers is a random sequence about the median. This is done using the *median* of the sequence of numbers. The process is explained in the next example.


University Corporation for Atmospheric Research

**EXAMPLE 5** 

## RUNS TEST ABOUT THE MEDIAN

Silver iodide seeding of summer clouds was done over the Santa Catalina mountains of Arizona. Of great importance is the direction of the wind during the seeding process. A sequence of consecutive days gave the following compass readings for wind direction at seeding level at 5 A.M. (0° represents true north) (Reference: *Proceedings of the National Academy of Science*, Vol. 68, pp. 649–652).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 174 | 160 | 175 | 288 | 195 | 140 | 124 | 219 | 197 | 184 |
| 183 | 224 | 33 | 49 | 175 | 74 | 103 | 166 | 27 | 302 |
| 61 | 72 | 93 | 172 | | | | | | |

We will test this sequence for randomness above and below the median using a 5% level of significance.

**Part I:** Adjust the sequence so that it has only two symbols, A and B.

**SOLUTION:** First rank-order the data and find the median (see Section 3.1). Doing this, we find the median to be 169. Next, give each data value in the original sequence the label A if it is *above* the median and the label B if it is *below* the median. Using the original sequence, we get

A | B | AAA | BB | AAAAA | BB | A | BBBB | A | BBB | A

We see that

$n_1 = 12$ (number of A's) $n_2 = 12$ (number of B's) $R = 11$ (number of runs)

*Note:* In this example, none of the data values actually equals the median. If a data value *equals the median*, we put neither A nor B in the sequence. This eliminates from the sequence any data values that equal the median.

**Part II:** Test the sequence of A and B symbols for randomness.

(a) What is the level of significance $\alpha$? State the null and alternate hypotheses.

**SOLUTION:** $\alpha = 0.05$

$H_0$: The symbols A and B are randomly mixed in the sequence.

$H_1$: The symbols A and B are not randomly mixed in the sequence.

(b) Find the sample test statistic $R$ and the parameters $n_1$ and $n_2$.

**SOLUTION:** As shown in Part I, for the sequence of A's and B's,

$n_1 = 12$; $n_2 = 12$; $R = 11$

(c) Use Table 10 of Appendix II to find the critical values $c_1$ and $c_2$.

**SOLUTION:** Since $n_1 = 12$ and $n_2 = 12$, we find $c_1 = 7$ and $c_2 = 19$.

(d) Conclude the test.

**SOLUTION:**

| $R \leq 7$ | ✓ $8 \leq R \leq 18$ | $R \geq 19$ |
|---|---|---|
| Reject $H_0$. | ✓ Fail to reject $H_0$. | Reject $H_0$. |

Since $R = 11$, we fail to reject $H_0$ at the 5% level of significance.

(e) *Interpret* the conclusion in the context of the problem.

**SOLUTION:** At the 5% level of significance, there is insufficient evidence to conclude that the sequence of wind directions above and below the median direction is not random.

**PROCEDURE**

## How to construct a runs test for randomness

***Setup***

You need a sequence (ordered set) consisting of two symbols. If your sequence consists of measurements of some type, then convert it to a sequence of two symbols in the following way:

(a) Find the median of the entries in the sequence.

(b) Label an entry A if it is above the median and B if it is below the median. If an entry equals the median, then put neither A nor B in the sequence.

Now you have a sequence with two symbols.

Let $n_1$ = number of times the first symbol occurs in the sequence.

$n_2$ = number of times the second symbol occurs in the sequence.

*Note:* Either symbol can be called the "first" symbol.

Let $R$ = number of runs in the sequence.

***Procedure***

1. The *level of significance is* $\alpha = 0.05$. The *null and alternate hypotheses* are:

   $H_0$: The two symbols are randomly mixed in the sequence.

   $H_1$: The two symbols are not randomly mixed in the sequence.

2. The *sample test statistic* is the number of runs $R$.
3. Use Table 10, Appendix II, with parameters $n_1$ and $n_2$ to find the *lower and upper critical values* $c_1$ and $c_2$.
4. Use the *critical values* $c_1$ and $c_2$ in the following *decision process*.

| $R \leq c_1$ | $c_1 + 1 \leq R \leq c_2 - 1$ | $R \leq c_2$ |
|---|---|---|
| Reject $H_0$. | Fail to reject $H_0$. | Reject $H_0$. |

5. *Interpret your conclusion* in the context of the application.

   *Note:* If your original sequence consisted of measurements (not just symbols), it is important to remember that you are testing for randomness about the median of these measurements. In any case, you are testing for randomness regarding a mix of two symbols in a given sequence.

Problem 11 describes how to use a normal approximation for the sample test statistic. Problem 12 gives additional practice.

COMMENT In many applications, $n_1 \leq 20$ and $n_2 \leq 20$. What happens if either $n_1 > 20$ or $n_2 > 20$? In this case, you can use the normal approximation, which is presented in Problems 11 and 12 at the end of this section.

**GUIDED EXERCISE 7** **Runs test for randomness of two symbols**

The majority party of the U.S. Senate for each year from 1973 to 2003 is shown below, where D and R represent Democrat and Republican, respectively (Reference: *Statistical Abstract of the United States*).

D D D D R R R D D D D R R R R D D R

Test the sequence for randomness. Use a 5% level of significance.

*Continued*

GUIDED EXERCISE 7 *continued*

(a) What is $\alpha$? State the null and alternate hypotheses.

$\alpha = 0.05$

$H_0$: The two symbols are randomly mixed.

$H_1$: The two symbols are not randomly mixed.

(b) Block the sequence into runs. Find the values of $n_1$, $n_2$, and $R$.

DDDD | RRR | DDDD | RRRR | DD | R

Letting D be the first symbol, we have

$n_1 = 10$; $n_2 = 8$; $R = 6$

(c) Use Table 10 of Appendix II to find the critical values $c_1$ and $c_2$.

Lower critical value $c_1 = 5$
Upper critical value $c_2 = 15$

(d) Using critical values, do you reject or fail to reject $H_0$?

| $R \le 5$ | ✓ $6 \le R \le 14$ | $R \ge 15$ |
|---|---|---|
| Reject $H_0$. | ✓ Fail to reject $H_0$. | Reject $H_0$. |

Since $R = 6$, we fail to reject $H_0$.

(e) ***Interpret*** the conclusion in the context of the application.

The sequence of party control of the U.S. Senate appears to be random. At the 5% level of significance, the evidence is insufficient to reject $H_0$, that the sequence is random.

## GUIDED EXERCISE 8 Runs test for randomness about the median

The national percentage distribution of burglaries is shown by month, starting in January (Reference: *FBI Crime Report,* U.S. Department of Justice).

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7.8 | 6.7 | 7.6 | 7.7 | 8.3 | 8.2 | 9.0 | 9.1 | 8.6 | 9.3 | 8.8 | 8.9 |

Test the sequence for randomness about the median. Use a 5% level of significance.

(a) What is $\alpha$? State the null and alternate hypotheses.

$\alpha = 0.05$

$H_0$: The sequence of values above and below the median is random.

$H_1$: The sequence of values above and below the median is not random.

(b) Find the median. Assign the symbol A to values above the median and the symbol B to values below the median. Next block the sequence of A's and B's into runs. Find $n_1$, $n_2$, and $R$.

First order the numbers. Then find the median. Median = 8.45. The original sequence translates to

BBBBBB | AAAAAA

$n_1 = 6$; $n_2 = 6$; $R = 2$

(c) Use Table 10 of Appendix II to find the critical values $c_1$ and $c_2$.

Lower critical value $c_1 = 3$
Upper critical value $c_2 = 11$

(d) Using the critical values, do you reject or fail to reject $H_0$?

| ✓ $R \le 3$ | $4 \le R \le 10$ | $R \le 11$ |
|---|---|---|
| ✓ Reject $H_0$. | Fail to reject $H_0$. | Reject $H_0$. |

Since $R = 2$, we reject $H_0$.

(e) ***Interpret*** the conclusion in the context of the application.

At the 5% level of significance, there is sufficient evidence to claim that the sequence of burglaries is not random about the median. It appears that from January to June, there tend to be fewer burglaries.

**TECH NOTES**

**Minitab** Enter your sequence of numbers in a column. Use the menu choices **Stat ➤ Nonparametrics ➤ Runs.** In the dialogue box, select the column containing the sequence. The default is to test the sequence for randomness above and below the mean. Otherwise, you can test for randomness above and below any other value, such as the median.

## SECTION 11.4 PROBLEMS

1. *Statistical Literacy* To apply a runs test for randomness as described in this section to a sequence of symbols, how many different symbols are required?

2. *Statistical Literacy* Suppose your data consist of a sequence of numbers. To apply a runs test for randomness about the median, what process do you use to convert the numbers into two distinct symbols?

For Problems 3–10, please provide the following information.

(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find the sample test statistic $R$, the number of runs.
(c) Find the upper and lower critical values in Table 10 of Appendix II.
(d) Conclude the test.
(e) *Interpret* the conclusion in the context of the application.

3. *Presidents: Party Affiliation* For each successive presidential term from Teddy Roosevelt to George W. Bush (first term), the party affiliation controlling the White House is shown below, where R designates Republican and D designates Democrat (Reference: *The New York Times Almanac*).

   R R R D D R R D D D D D R D R R D R R R D D R

   *Historical Note:* In cases in which a president died in office or resigned, the period during which the vice president finished the term is not counted as a new term. Test the sequence for randomness. Use $\alpha = 0.05$.

4. *Congress: Party Affiliation* The majority party of the U.S. House of Representatives for each year from 1973 to 2003 is shown below, where D and R represent Democrat and Republican, respectively (Reference: *Statistical Abstract of the United States*).

   D D D D D D D D D D D R R R R R R R

   Test the sequence for randomness. Use $\alpha = 0.05$.

5. *Cloud Seeding: Arizona* Researchers experimenting with cloud seeding in Arizona want a random sequence of days for their experiments (Reference: *Proceedings of the National Academy of Science*, Vol. 68, pp. 649–652). Suppose they have the following itinerary for consecutive days, where S indicates a day for cloud seeding and N indicates a day for no cloud seeding.

   S S S N S N S S S S N N S N S S S N N S S S S

   Test this sequence for randomness. Use $\alpha = 0.05$.

6. *Astronomy: Earth's Rotation* Changes in the earth's rotation are exceedingly small. However, a very long-term trend could be important. (Reference: *Journal of Astronomy*, Vol. 57, pp. 125–146). Let I represent an increase and D a decrease in the rate of the earth's rotation. The following sequence represents historical increases and decreases measured every consecutive fifth year.

   D D D D D I I I D D D D D I I I I I I I I I D I I I I

   Test the sequence for randomness. Use $\alpha = 0.05$.

7. ***Random Walk: Stocks*** Many economists and financial experts claim that the *price level* of a stock or bond is not random; rather, the *price changes* tend to follow a random sequence over time. The following data represent annual percentage returns on Vanguard Total Stock Index for a sequence of recent years. This fund represents nearly all publicly traded U.S. stocks (Reference: *Morningstar Mutual Fund Analysis*).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 10.4 | 10.6 | −0.2 | 35.8 | 21.0 | 31.0 | 23.3 | 23.8 | −10.6 |
| −11.0 | −21.0 | 12.8 | | | | | | |

(i) Convert this sequence of numbers to a sequence of symbols A and B, where A indicates a value above the median and B a value below the median.
(ii) Test the sequence for randomness about the median. Use $\alpha = 0.05$.

8. ***Random Walk: Bonds*** The following data represent annual percentage returns on Vanguard Total Bond Index for a sequence of recent years. This fund represents nearly all publicly traded U.S. bonds (Reference: *Morningstar Mutual Fund Analysis*).

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7.1 | 9.7 | −2.7 | 18.2 | 3.6 | 9.4 | 8.6 | −0.8 | 11.4 | 8.4 | 8.3 | 0.8 |

(i) Convert this sequence of numbers to a sequence of symbols A and B, where A indicates a value above the median and B a value below the median.
(ii) Test the sequence for randomness about the median. Use $\alpha = 0.05$.

9. ***Civil Engineering: Soil Profiles*** Sand and clay studies were conducted at the West Side Field Station of the University of California (Reference: Professor D. R. Nielsen, University of California, Davis). Twelve consecutive depths, each about 15 cm deep, were studied and the following percentages of sand in the soil were recorded.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 19.0 | 27.0 | 30.0 | 24.3 | 33.2 | 27.5 | 24.2 | 18.0 | 16.2 | 8.3 | 1.0 | 0.0 |

(i) Convert this sequence of numbers to a sequence of symbols A and B, where A indicates a value above the median and B a value below the median.
(ii) Test the sequence for randomness about the median. Use $\alpha = 0.05$.

10. ***Civil Engineering: Soil Profiles*** Sand and clay studies were conducted at the West Side Field Station of the University of California (Reference: Professor D. R. Nielsen, University of California, Davis). Twelve consecutive depths, each about 15 cm deep, were studied and the following percentages of clay in the soil were recorded.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 47.4 | 43.4 | 48.4 | 42.6 | 41.4 | 40.7 | 46.4 | 44.8 | 36.5 | 35.7 | 33.7 | 42.6 |

(i) Convert this sequence of numbers to a sequence of symbols A and B, where A indicates a value above the median and B a value below the median.
(ii) Test the sequence for randomness about the median. Use $\alpha = 0.05$.

11. ***Expand Your Knowledge: Either $n_1 > 20$ or $n_2 > 20$*** For each successive presidential term from Franklin Pierce (the 14th president, elected in 1853) to George W. Bush (43rd president), the party affiliation controlling the White House is shown below, where R designates Republican and D designates Democrat (Reference: *The New York Times Almanac*).

*Historical Note:* We start this sequence with the 14th president because earlier presidents belonged to political parties such as the Federalist or Wigg (not Democratic or Republican) party. In cases in which a president died in office or resigned, the period during which the vice president finished the term is not counted as a new term. The one exception is the case in which Lincoln (a Republican) was assassinated and the vice president Johnson (a Democrat) finished the term.

D D R R D R R R R D R D R R R R D D R R

D D D D D R R D D R R D R R R D D R

Test the sequence for randomness at the 5% level of significance. Use the following outline.

(a) State the null and alternate hypotheses.

(b) Find the number of runs $R$, $n_1$, and $n_2$. Let $n_1 =$ number of Republicans and $n_2 =$ number of Democrats.

(c) In this case, $n_1 = 21$, so we cannot use Table 10 of Appendix II to find the critical values. Whenever either $n_1$ or $n_2$ exceeds 20, the number of runs $R$ has a distribution that is approximately normal, with

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

We convert the number of runs $R$ to a $z$ value, and then use the normal distribution to find the critical values. Convert the sample test statistic $R$ to $z$ using the formula

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

(d) The critical values of a normal distribution for a two-tailed test with level of significance $\alpha = 0.05$ are $-1.96$ and $1.96$ (see Table 5(c) of Appendix II). Reject $H_0$ if the sample test statistic $z \leq -1.96$ or if the sample test statistic $z \geq 1.96$. Otherwise, do not reject $H_0$.

| Sample $z \leq -1.96$ | $-1.96 <$ sample $z < 1.96$ | Sample $z \geq 1.96$ |
|---|---|---|
| Reject $H_0$. | Fail to reject $H_0$. | Reject $H_0$. |

Using this decision process, do you reject or fail to reject $H_0$ at the 5% level of significance? What is the $P$-value for this two-tailed test? At the 5% level of significance, do you reach the same conclusion using the $P$-value that you reach using critical values? Explain.

(e) ***Interpret*** your results in the context of the application.

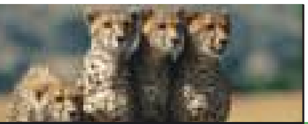
12. ***Expand Your Knowledge: Either $n_1 > 20$ or $n_2 > 20$*** Professor Cornish studied rainfall cycles and sunspot cycles (Reference: *Australian Journal of Physics*, Vol. 7, pp. 334–346). Part of the data include amount of rain (in mm) for 6-day intervals. The following data give rain amounts for consecutive 6-day intervals at Adelaide, South Australia.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 29 | 6 | 0 | 68 | 0 | 0 | 2 | 23 | 5 | 18 | 0 | 50 | 163 |
| 64 | 72 | 26 | 0 | 0 | 3 | 8 | 142 | 108 | 3 | 90 | 43 | 2 | 5 |
| 0 | 21 | 2 | 57 | 117 | 51 | 3 | 157 | 43 | 20 | 14 | 40 | 0 | 23 |
| 18 | 73 | 25 | 64 | 114 | 38 | 31 | 72 | 54 | 38 | 9 | 1 | 17 | 0 |
| 13 | 6 | 2 | 0 | 1 | 5 | 9 | 11 | | | | | | |

Verify that the median is 17.5.

(a) Convert this sequence of numbers to a sequence of symbols A and B, where A indicates a value above the median and B a value below the median.

(b) Test the sequence for randomness about the median at the 5% level of significance. Use the large sample theory outlined in Problem 11.



# Chapter Review

## SUMMARY

When we cannot assume that data come from a normal, binomial, or Student's $t$ distribution, we can employ tests that make no assumptions about data distribution. Such tests are called nonparametric tests. We studied four widely used tests: the sign test, the rank-sum test, the Spearman rank correlation coefficient test, and the runs test for randomness. Nonparametric tests have both advantages and disadvantages:

- Advantages of nonparametric tests

  No requirements concerning the distributions of populations under investigation.

  Easy to use.

- Disadvantages of nonparametric tests

  Waste information.

  Are less sensitive.

It is usually good advice to use standard tests when possible, keeping nonparametric tests for situations wherein assumptions about the data distribution cannot be made.

## IMPORTANT WORDS & SYMBOLS

**Section 11.1**
Nonparametric statistics 678
Sign test 678

**Section 11.2**
Rank-sum test 686

**Section 11.3**
Monotone relationship 695
Spearman rank correlation coefficient $r_s$ 696
Population Spearman rank correlation coefficient $\rho_s$ 696

**Section 11.4**
Sequence 706
Run 706
Runs test for randomness 705

## VIEWPOINT Lending a Hand

*Whom would you ask for help if you were sick? in need of money? upset with your spouse? depressed? Consider the following claims: People look to sisters for emotional help and brothers for physical help. After that, people look to parents, clergy, or friends. Can you think of nonparametric tests to study such claims? For more information, see* American Demographics, *Vol. 18, No. 8.*

## CHAPTER REVIEW PROBLEMS

1. *Statistical Literacy* For nonparametric tests, what assumptions, if any, need to be made concerning the distributions of the populations under investigation?
2. *Critical Thinking* Suppose you want to test whether there is a difference in means in a matched pair, "before and after" situation. If you know that the populations under investigation are at least mound-shaped and symmetrical and you have a large sample, is it better to use the parametric paired differences test or the nonparametric sign test for matched pairs? Explain.

For Problems 3–10, please provide the following information.
(a) State the test used.
(b) Give $\alpha$. State the null and alternate hypotheses.
(c) Find the sample test statistic.

(d) For the sign test, rank-sum test, and Spearman correlation coefficient test, find the $P$-value of the sample test statistic. For the runs test of randomness, find the critical values from Table 10 of Appendix II.

(e) Conclude the test and ***interpret*** the results in the context of the application.

3. ***Chemistry: Lubricant*** In the production of synthetic motor lubricant from coal, a new catalyst has been discovered that seems to affect the viscosity index of the lubricant. In an experiment consisting of 23 production runs, 11 used the new catalyst and 12 did not. After each production run, the viscosity index of the lubricant was determined to be as follows.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| With catalyst | 1.6 | 3.2 | 2.9 | 4.4 | 3.7 | 2.5 | 1.1 | 1.8 | 3.8 | 4.2 | 4.1 | |
| Without catalyst | 3.9 | 4.6 | 1.5 | 2.2 | 2.8 | 3.6 | 2.4 | 3.3 | 1.9 | 4.0 | 3.5 | 3.1 |

The two samples are independent. Use a 0.05 level of significance to test the null hypothesis that the viscosity index is unchanged by the catalyst against the alternate hypothesis that the viscosity index has changed.

4. ***Self-Improvement: Memory*** Professor Adams wrote a book called *Improving Your Memory*. The professor claims that if you follow the program outlined in the book, your memory will definitely improve. Fifteen people took the professor's course, in which the book and its program were used. On the first day of class, everyone took a memory exam; and on the last day, everyone took a similar exam. The paired scores for each person follow.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Last exam | 225 | 120 | 115 | 275 | 85 | 76 | 114 | 200 | 99 | 135 | 170 | 110 | 216 | 280 | 78 |
| First exam | 175 | 110 | 115 | 200 | 60 | 85 | 160 | 190 | 70 | 110 | 140 | 10 | 190 | 200 | 92 |

Use a 0.05 level of significance to test the null hypothesis that the scores are the same whether or not people have taken the course against the alternate hypothesis that the scores of people who have taken the course are higher.

5. ***Sales: Paint*** A chain of hardware stores is trying to sell more paint by mailing pamphlets describing the paint. In 15 communities containing one of these hardware stores, the paint sales (in dollars) were recorded for the months before and after the ads were sent out. The paired results for each store follow.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sales after | 610 | 150 | 790 | 288 | 715 | 465 | 280 | 640 | 500 | 118 | 265 | 365 | 93 | 217 | 280 |
| Sales before | 460 | 216 | 640 | 250 | 685 | 430 | 220 | 470 | 370 | 118 | 117 | 360 | 93 | 291 | 430 |

Use a 0.01 level of significance to test the null hypothesis that the advertising had no effect on sales against the alternate hypothesis that it improved sales.

6. ***Dogs: Obedience School*** An obedience school for dogs experimented with two methods of training. One method involved rewards (food, praise); the other involved no rewards. The dogs were randomly placed into two independent groups of 11 each. The number of sessions required to train each of 22 dogs follows.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| With rewards | 12 | 17 | 15 | 10 | 16 | 20 | 9 | 23 | 8 | 14 | 10 |
| No rewards | 19 | 22 | 11 | 18 | 13 | 25 | 24 | 28 | 21 | 20 | 21 |

Use a 0.05 level of significance to test the hypothesis that the number of sessions was the same for the two groups against the alternate hypothesis that the number of sessions was not the same.

7. *Training Program: Fast Food* At McDouglas Hamburger stands, each employee must undergo a training program before he or she is assigned. A group of nine people went through the training program and were assigned to work at the Teton Park McDouglas Hamburger stand. Rankings in performance after the training program and after one month on the job are shown (a rank of 1 is for best performance).

| Employee | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Rank, training program | 8 | 9 | 7 | 3 | 6 | 4 | 1 | 2 | 5 |
| Rank on job | 9 | 8 | 6 | 7 | 5 | 1 | 3 | 4 | 2 |

Using a 0.05 level of significance, test the claim that there is a monotone-increasing relation between rank from the training program and rank in performance on the job.

8. *Cooking School: Chocolate Mousse* Two expert French chefs judged chocolate mousse made by students in a Paris cooking school. Each chef ranked the best chocolate mousse as 1.

| Student | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Rank by Chef Pierre | 4 | 2 | 3 | 1 | 5 |
| Rank by Chef André | 4 | 1 | 2 | 3 | 5 |

Use a 0.10 level of significance to test the claim that there is a monotone relation (either way) between ranks given by Chef Pierre and by Chef André.

9. *Education: True–False Questions* Dr. Gill wants to arrange the answers to a true–false exam in random order. The answers in order of occurrence are shown below.

T T T T F T T F F T T T T F F F F F F T T T T T T

Test the sequence for randomness using $\alpha = 0.05$.

10. *Agriculture: Wheat* For the past 16 years, the yields of wheat (in tons) grown on a plot at Rothamsted Experimental Station (England) are shown below. The sequence is by year.

3.8 1.9 0.6 1.7 2.0 3.5 3.0 1.4 2.7 2.3 2.6 2.1
2.4 2.7 1.8 1.9

Use level of significance 5% to test for randomness about the median.

## DATA HIGHLIGHTS: GROUP PROJECTS

Break into small groups and discuss the following topics. Organize a brief outline in which you summarize the main points of your group discussion.

In the world of business and economics, to what extent do assets determine profits? Do the big companies with large assets always make more profits? Is there a rank correlation between assets and profits? The following table is based on information taken from

| Company | Asset Rank | Profit Rank |
|---|---|---|
| Pepsico | 4 | 2 |
| McDonald's | 1 | 1 |
| Aramark | 6 | 4 |
| Darden Restaurants | 7 | 5 |
| Flagstar | 11 | 11 |
| VIAD | 10 | 8 |
| Wendy's International | 2 | 3 |
| Host Marriott Services | 9 | 10 |
| Brinker International | 5 | 7 |
| Shoney's | 3 | 6 |
| Food Maker | 8 | 9 |

*Fortune* (Vol. 135, No. 8). A rank of 1 means highest profits or highest assets. The companies are food service companies.

(a) Compute the Spearman rank correlation coefficient for these data.

(b) Using a 5% level of significance, test the claim that there is a monotone-increasing relation between the ranks of earnings and growth.

(c) Decide whether you should reject or not reject the null hypothesis. Interpret your conclusion in the context of the problem.

(d) As an investor, what are some other features of food companies that you might be interested in ranking? Identify any such features that you think might have a monotone relation.

## LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. (a) What do we mean by the term *nonparametric statistics*? What do we mean by the term *parametric statistics*? How do nonparametric methods differ from the methods we studied earlier?

   (b) What are the advantages of nonparametric statistical methods? How can they be used in problems to which other methods we have learned would not apply?

   (c) Are there disadvantages to nonparametric statistical methods? What do we mean when we say that nonparametric methods tend to waste information? Why do we say that nonparametric methods are not as *sensitive* as parametric methods?

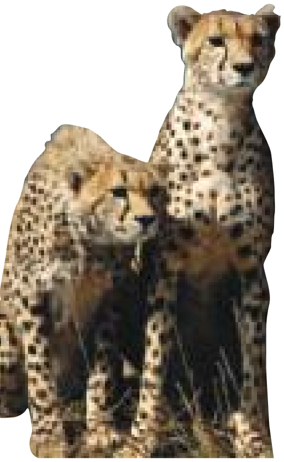
   (d) List three random variables from ordinary experience to which you think nonparametric methods would definitely apply and the application of parametric methods would be questionable.

2. Outline the basic logic and ideas behind the sign test. Describe how the binomial probability distribution was used in the construction of the sign test. What assumptions must be made about the sign test? Why is the sign test so extremely general in its possible applications? Why is it a special test for "before and after" studies?

3. Outline the basic logic and ideas behind the rank-sum test. Under what conditions would you use the rank-sum test and *not* the sign test? What assumptions must be made in order to use the rank-sum test? List two advantages the rank-sum test has that the methods of Section 8.5 do not have. List some advantages the methods of Section 8.5 have that the rank-sum test does not have.

4. What do we mean by a monotone relationship between two variables $x$ and $y$? What do we mean by ranked variables? Give a graphic example of two variables $x$ and $y$ that have a monotone relationship but do *not* have a linear relationship. Does the Spearman test check for a monotone relationship or a linear relationship? Under what conditions does the Pearson product-moment correlation coefficient reduce to the Spearman rank correlation coefficient? Summarize the basic logic and ideas behind the test for Spearman rank correlation. List variables $x$ and $y$ from daily experience for which you think a strong Spearman rank correlation coefficient exists even though the variables are *not* linearly related.

5. What do we mean by a runs test for randomness? What is a run in a sequence? How can we test for randomness about the median? Why is this an important concept? List at least three applications from your own experience.



# Cumulative Review Problems

## CHAPTERS 10–11

1. *Goodness-of-Fit Test: Rare Events* This cumulative review problem uses material from Chapters 3, 5, and 10. Recall that the Poisson distribution deals with rare events. Death from the kick of a horse is a rare event, even in the Prussian army. The following data are a classic example of a Poisson application to rare events. A reproduction of the original data can be found in C. P. Winsor, *Human Biology,* Vol. 19, pp. 154–161. The data represent the number of deaths from the kick of a horse per army corps per year for 10 Prussian army corps for 20 years (1875–1894). Let $x$ represent the number of deaths and $f$ the frequency of $x$ deaths.

| $x$ | 0 | 1 | 2 | 3 or more |
|---|---|---|---|---|
| $f$ | 109 | 65 | 22 | 4 |

(a) First, we fit the data to a Poisson distribution (see Section 5.4).

Poission distribution: $P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

where $\lambda \approx \bar{x}$ (sample mean of $x$ values)
From our study of weighted averages (see Section 3.1),

$$\bar{x} = \frac{\Sigma xf}{\Sigma f}$$

Verify that $\bar{x} \approx 0.61$ *Hint:* For the category 3 or more, use 3.

(b) Now we have $P(x) = \dfrac{e^{-0.61}(0.61)^x}{x!}$ for $x = 0, 1, 2, 3. \ldots$

Find $P(0)$, $P(1)$, $P(2)$, and $P(3 \le x)$. Round to three places after the decimal.

(c) The total number of observations is $\Sigma f = 200$. For a given $x$, the expected frequency of $x$ deaths is $200P(x)$. The following table gives the observed frequencies $O$ and the expected frequencies $E = 200P(x)$.

| $x$ | $O = f$ | $E = 200P(x)$ |
|---|---|---|
| 0 | 109 | $200(0.543) = 108.6$ |
| 1 | 65 | $200(0.331) = 66.2$ |
| 2 | 22 | $200(0.101) = 20.2$ |
| 3 or more | 4 | $200(0.025) = 5$ |

Compute $\chi^2 = \Sigma \dfrac{(O - E)^2}{E}$

(d) State the null and alternate hypotheses for a chi-square goodness-of-fit test. Set the level of significance to be $\alpha = 0.01$. Find the $P$-value for a goodness-of-fit test. Interpret your conclusion in the context of this application. Is there reason to believe that the Poisson distribution fits the raw data provided by the Prussian army? Explain.

2. *Test of Independence: Agriculture* Three types of fertilizer were used on 132 identical plots of maize. Each plot was harvested and the yield (in kg) was recorded (Reference: Caribbean Agricultural Research and Development Institute).

| | Type of Fertilizer | | | |
|---|---|---|---|---|
| Yield (kg) | I | II | III | Row Total |
| 0–2.9 | 12 | 10 | 15 | 37 |
| 3.0–5.9 | 18 | 21 | 11 | 50 |
| 6.0–8.9 | 16 | 19 | 10 | 45 |
| Column Total | 46 | 50 | 36 | 132 |

Use a 5% level of significance to test the hypothesis that type of fertilizer and yield of maize are independent. Interpret the results.

3. *Testing and Estimating Variances: Iris* Random samples of two species of iris gave the following petal lengths (in cm) (Reference: R. A. Fisher, *Annals of Eugenics*, Vol. 7).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$, *Iris virginica* | 5.1 | 5.9 | 4.5 | 4.9 | 5.7 | 4.8 | 5.8 | 6.4 | 5.6 | 5.9 |
| $x_2$, *Iris versicolor* | 4.5 | 4.8 | 4.7 | 5.0 | 3.8 | 5.1 | 4.4 | 4.2 | | |

(a) Use a 5% level of significance to test the claim that the population standard deviation of $x_1$ is larger than 0.55.
(b) Find a 90% confidence interval for the population standard deviation of $x_1$.
(c) Use a 1% level of significance to test the claim that the population variance of $x_1$ is larger than that of $x_2$. Interpret the results.

4. *Sign Test: Wind Direction* The following data are paired by date. Let $x$ and $y$ be random variables representing wind direction at 5 A.M. and 5 P.M., respectively (units are degrees on a compass, with 0° representing true north). The readings were taken at seeding level in a cloud seeding experiment. (Reference: *Proceedings of the National Academy of Science,* Vol. 68, pp. 649–652.) A random sample of days gave the following information.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $x$ | 177 | 140 | 197 | 224 | 49 | 175 | 257 | 72 | 172 |
| $y$ | 142 | 142 | 217 | 125 | 53 | 245 | 218 | 35 | 147 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $x$ | 214 | 265 | 110 | 193 | 180 | 190 | 94 | 8 | 93 |
| $y$ | 205 | 218 | 100 | 170 | 245 | 117 | 140 | 99 | 60 |

Use the sign test with a 5% level of significance to test the claim that the distributions of wind directions at 5 A.M. and 5 P.M. are different. Interpret the results.

5. *Rank-Sum Test: Apple Trees* Commercial apple trees usually consist of two parts grafted together. The upper part, or graft, determines the character of the fruit, while the root stock determines the size of the tree. (Reference: East Malling Research Station, England.) The following data are from two root stocks A and B. The data represent total extension growth (in meters) of the grafts after 4 years.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Stock A** | 2.81 | 2.26 | 1.94 | 2.37 | 3.11 | 2.58 | 2.74 | 2.10 | 3.41 | 2.94 | 2.88 |
| **Stock B** | **2.52** | 3.02 | 2.86 | 2.91 | 2.78 | 2.71 | 1.96 | 2.44 | 2.13 | 1.58 | 2.77 |

Use a 1% level of significance and the rank-sum test to test the claim that the distributions of growths are different for root stocks A and B. Interpret the results.


Eastcott-Momatiuk/The Image Works

6. *Spearman Rank Correlation: Calcium Tests* Random collections of nine different solutions of a calcium compound were given to two laboratories A and B. Each laboratory measured the calcium content (in mmol. per liter) and reported the results. The data are paired by calcium compound (Reference: *Journal of Clinical Chemistry and Clinical Biochemistry,* Vol. 19, pp. 395–426).

| Compound | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Lab A | 13.33 | 15.79 | 14.78 | 11.29 | 12.59 | 9.65 | 8.69 | 10.06 | 11.58 |
| Lab B | 13.17 | 15.72 | 14.66 | 11.47 | 12.65 | 9.60 | 8.75 | 10.25 | 11.56 |

(a) Rank-order the data using 1 for the lowest calcium reading. Make a table of ranks to be used in a Spearman rank correlation test.
(b) Use a 5% level of significance to test for a monotone relation (either way) between ranks. Interpret the results.

7. *Runs Test for Randomness: Sunspots* The January mean number of sunspots is recorded for a sequence of recent Januaries (Reference: *International Astronomical Union Quarterly Bulletin on Solar Activity*).

| | | | | | |
|---|---|---|---|---|---|
| 57.9 | 38.7 | 19.8 | 15.3 | 17.5 | 28.2 |
| 110.9 | 121.8 | 104.4 | 111.5 | 9.13 | 61.5 |
| 43.4 | 27.6 | 18.9 | 8.1 | 16.4 | 51.9 |

Use level of significance 5% to test for randomness about the median. Interpret the results.