10

Inferences About Differences

"Girl with Black Eye" by Norman Rockwell (1894–1978)

Norman Rockwell painted everyday people and situations. In this cover for the *Saturday Evening Post* (May 23, 1953), a young lady is about to have a conference with her school principal. So what!

"So what!"

—Anonymous

We have all heard the exclamation, "So what!" Philologists (people who study cultural linguistics) tell us that this expression is a shortened version of "So what is the difference!" They also tell us that there are similar popular or slang expressions about differences in all languages and cultures. It is human nature to challenge the claim that something is better, worse, or just simply different. In this chapter, we will focus on this very human theme by studying a variety of topics regarding questions of whether or not differences exist between two populations.

PREVIEW QUESTIONS

◇ What are the statistical advantages of paired data values? How do we construct statistical tests? (SECTION 10.1)

◇ How do we compare means from two independent populations when we know σ for each population? (SECTION 10.2)

◇ What if we want to compare means from two independent populations, but we do not know σ for each population? (SECTION 10.2)

◇ How do we use sample data to compare proportions from two independent populations? (SECTION 10.3)

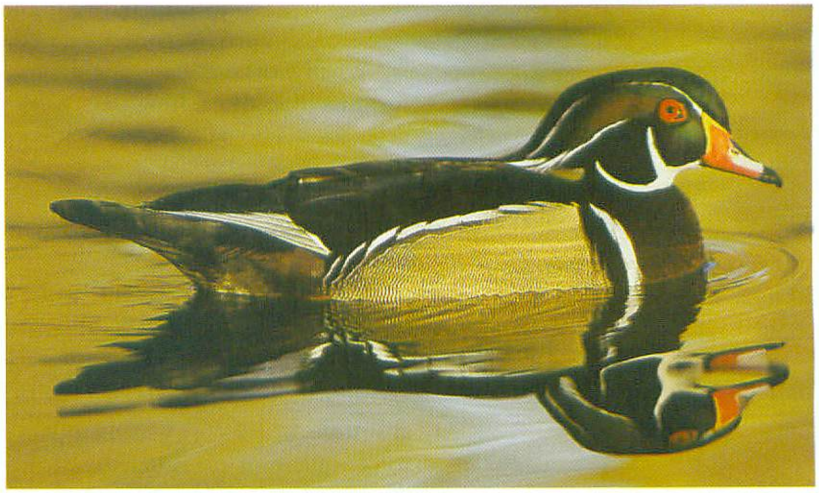
For on-line student resources, visit **math.college.hmco.com/students** and follow the Statistics links to the Brase/Brase, *Understanding Basic Statistics,* 4th edition web site.

FOCUS PROBLEM

The Trouble with Wood Ducks

The National Wildlife Federation published an article entitled "The Trouble with Wood Ducks" (*National Wildlife*, Vol. 31, No. 5). In this article, wood ducks are described as beautiful birds living in forested areas such as the Pacific Northwest and Southeast United States. Because of overhunting and habitat destruction, these birds were in danger of extinction. A federal ban on hunting wood ducks in 1918 helped save the species from extinction. Wood ducks like to nest in tree cavities. However, many such trees were disappearing due to heavy timber cutting. For a period of time it seemed that nesting boxes were the solution to disappearing trees. At first, the wood duck population grew, but after a few seasons, the population declined sharply. Good biology research combined with good statistics provided an answer to this disturbing phenomenon.

Cornell University professors of ecology Paul Sherman and Brad Semel found that the nesting boxes were placed too close to each other. Female wood ducks prefer a secluded nest that is a considerable distance from the next wood duck nest. In fact, female wood duck behavior changed when the nests were too close to each other. Some females would lay their eggs in another female's nest. The result was too many eggs in one nest. The biologists found that if there were too many eggs in a nest, the proportion of eggs that hatched was considerably reduced. In the long run, this meant a decline in the population of wood ducks.



In their study, Sherman and Semel used two placements of nesting boxes. Group I boxes were well separated from each other and well hidden by available brush. Group II boxes were highly visible and grouped closely together.

In group I boxes, there were a total of 474 eggs, of which a field count showed that about 270 hatched. In group II boxes, there were a total of 805 eggs, of which a field count showed that, again, about 270 hatched.

(a) Find a 95% confidence interval for $p_1 - p_2$. Does the interval indicate that the proportion of eggs hatched from group I nest box placements is higher than, lower than, or not different from the proportion of eggs hatched from group II nest boxes?

(b) Use a 5% level of significance to test the hypothesis that the proportion of hatches from group I placements is greater than that from group II placements. Do we reject or fail to reject H_0? Is this result consistent with the information from the 95% confidence interval?

(c) What conclusions about placement of nest boxes can be drawn? In the article, additional concerns are raised about the higher cost of placing and maintaining group I nest boxes. Also at issue is the cost efficiency per successful wood duck hatch. Data in the article do not include information that would help us answer questions of *cost* efficiency. However, the data presented do help us answer questions about the proportions of successful hatches in the two nest box configurations.

(See Problem 16 of Section 10.3.)

10.1 Tests Involving Paired Differences (Dependent Samples)

FOCUS POINTS

- Identify paired data and dependent samples.
- Explain the advantages of paired data tests.
- Compute differences and the sample test statistic.
- Estimate the *P*-value and conclude the test.

Many statistical applications use *paired data* samples to draw conclusions about the difference between two population means. Data *pairs* occur very naturally in "before and after" situations, where the *same* object or item is measured both before and after a treatment. Applied problems in social science, natural science, and business administration frequently involve a study of matching pairs. Psychological studies of identical twins; biological studies of plant growth on plots of land matched for soil type, moisture, and sun; and business studies on sales of matched inventories are examples of paired data studies.

Creating data pairs

When working with paired data, it is very important to have a definite and uniform method of creating data pairs that clearly utilizes a natural matching of characteristics. The next example and exercise demonstrate this feature.

EXAMPLE 1 *Paired data*

A shoe manufacturer claims that among the general population of adults in the United States, the average length of the left foot is longer than that of the right. To compare the average length of the left foot with that of the right, we can take a random sample of 15 U.S. adults and measure the length of the left foot and then the length of the right foot for each person in the sample. Is there a natural way of pairing the measurements? How many pairs will we have?

SOLUTION: In this case, we can pair each left foot measurement with the same person's right foot measurement. The person serves as the "matching link" between the two distributions. We will have 15 pairs of measurements. ◇

GUIDED EXERCISE 1

Paired data

A psychologist has developed a series of exercises called the Instrumental Enrichment (IE) Program, which he claims is useful in overcoming cognitive deficiencies in mentally retarded children. To test the program, extensive statistical tests are being conducted. In one experiment, a random sample of 10-year-old students with IQ scores below 80 was selected. An IQ test was given to these students before they spent 2 years in an IE Program, and an IQ test was given to the same students after the program.

(a) On what basis can you pair the IQ scores? → Take the "before and after" IQ scores of each individual student.

(b) If there were 20 students in the sample, how many data pairs would you have? → Twenty data pairs. Note that there would be 40 IQ scores, but only 20 pairs.

◇ **COMMENT** To compare two populations, we cannot always employ paired data tests, but when we can, what are the advantages? Using matched or paired data often can reduce the danger of introducing extraneous or uncontrollable factors into our sample measurements, because the matched or paired data have essentially the *same* characteristics except for the *one* characteristic that is being measured. Furthermore, it can be shown that pairing data has the theoretical effect of reducing measurement variability (i.e., variance), which increases the precision of statistical conclusions. ◇

Two samples are **dependent** if each data value in one sample can be paired with a corresponding data value in the other sample.

When we wish to compare the means of two samples, the first item to be determined is whether or not there is a natural pairing between the data in the two samples. Again, data pairs are created from "before and after" situations, or from matching data by using studies of the same object, or by a process of taking measurements of closely matched items.

Testing the differences d

When testing *paired* data, we take the differences d of the data pairs *first* and look at the mean difference $\bar{d}$. Then we use a test on $\bar{d}$. Theorem 10.1 provides the basis for our work with paired data.

◇ **THEOREM 10.1** Consider a random sample of n data pairs. Suppose the differences d between the first and second members of each data pair are (approximately) normally distributed with population mean μ_d. Then the t values

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the sample mean of the d values, n is the number of data pairs, and

$$s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}}$$

is the sample standard deviation of the d values, follow a Student's t distribution with degrees of freedom $d.f. = n - 1$. ◇

Hypotheses for testing the mean of paired differences

When testing the mean of the differences of paired data values, the null hypothesis is that there is no difference among the pairs. That is, the mean of the differences μ_d is zero.

$$H_0: \mu_d = 0$$

The alternate hypothesis depends on the problem and can be

$H_1: \mu_d < 0$	$H_1: \mu_d > 0$	$H_1: \mu_d \neq 0$
(left-tailed)	(right-tailed)	(two-tailed)

Sample test statistic

For paired difference tests, we make our decision regarding H_0 according to the evidence of the sample mean $\bar{d}$ of the differences of measurements. By Theorem 10.1, we convert the sample test statistic $\bar{d}$ to a t value using the formula

$$t = \frac{\bar{d} - \mu_d}{(s_d/\sqrt{n})} \text{ with } d.f. = n - 1$$

where s_d = sample standard deviation of the differences d

n = number of data pairs

μ_d = 0 as specified in H_0

P-values from Table 4 of the Appendix

To find the P-value (or an interval containing the P-value) corresponding to the test statistic t computed from $\bar{d}$, we use the Student's t distribution table (Table 4 of the Appendix). Recall from Section 9.2 that we find the test statistic t (or, if t is negative, $|t|$) in the row headed by $d.f. = n - 1$, where n is the number of data pairs. The P-value for the test statistic is the column entry in the *one-tail area* row for one-tailed tests (right or left). For two-tailed tests, the P-value is the column entry in the *two-tail area* row. Usually the exact test statistic t is not in the table, so we obtain an interval that contains the P-value by using adjacent entries in the table. Table 10-1 gives the basic structure for using the Student's t distribution table to find the P-value or an interval containing the P-value.

TABLE 10-1 Using Student's *t* Distribution Table for *P*-values

For one-tailed tests:	one-tail area	*P*-value	*P*-value
For two-tailed tests:	two-tail area	*P*-value	*P*-value
Use row header	$d.f. = n - 1$	↑ Find t value	

With the preceding information, you are now ready to test paired differences. First let's summarize the procedure.

PROCEDURE

How to test paired differences using the Student's t distribution

Obtain a simple random sample of n matched data pairs A, B. Let d be a random variable representing the difference between the values in a matched data pair. Compute the sample mean $\bar{d}$ and sample standard deviation s_d.

If you can assume that d has a normal distribution or simply has a mound-shaped symmetric distribution, then any sample size n will work. If you cannot assume this, then use a sample size $n \geq 30$.

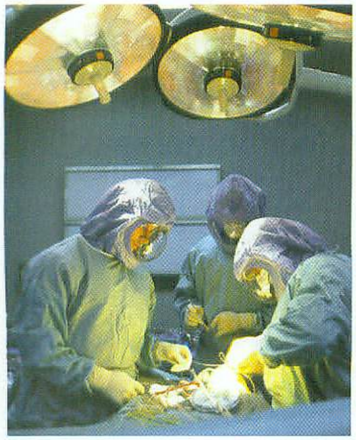
1. Use the *null hypothesis* of *no difference*, H_0: $\mu_d = 0$. In the context of the application, choose the *alternate hypothesis* to be H_1: $\mu_d > 0$, or $\mu_d < 0$, or $\mu_d \neq 0$. Set the *level of significance* α.
2. Use $\bar{d}$, s_d, the sample size n, and $\mu_d = 0$ from the null hypothesis to compute the sample *test statistic*
$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d}\sqrt{n}}{s_d}$$
with degrees of freedom $d.f. = n - 1$.
3. Use the Student's t distribution and the type of test, one-tailed or two-tailed, to find (or estimate) 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. *State your conclusion* in the context of the application.

EXAMPLE 2
Paired difference test

A team of heart surgeons at Saint Ann's Hospital knows that many patients who undergo corrective heart surgery have a dangerous buildup of anxiety before their scheduled operations. The staff psychiatrist at the hospital has started a new counseling program intended to reduce this anxiety. A test of anxiety is given to patients who know they must undergo heart surgery. Then each patient participates in a series of counseling sessions with the staff psychiatrist. At the end of the counseling sessions, each patient is retested to determine anxiety level. Table 10-2 indicates the results for a random sample of nine patients. Higher scores mean higher levels of anxiety.

TABLE 10-2

Patient	B Score before Counseling	A Score after Counseling	$d = B - A$ Difference
Jan	121	76	45
Tom	93	93	0
Diane	105	64	41
Kamal	115	117	−2
Mike	130	82	48
Bill	98	80	18
Enrique	142	79	63
Carol	118	67	51
Alice	125	89	36



From the given data, can we conclude that the counseling sessions reduce anxiety? Use a 0.01 level of significance.

SOLUTION: Before we answer this question, let us notice two important points: (1) we have a *random sample* of nine patients, and (2) we have a *pair* of measurements taken on the same patient before and after counseling sessions. In our problem, the sample size is $n = 9$ pairs (i.e., patients), and the d values are found in the fourth column of Table 10-2 on the previous page.

(a) Note the level of significance and set the hypotheses.
In the problem statement, $\alpha = 0.01$. We want to test the claim that the counseling sessions reduce anxiety. This means that the anxiety level before counseling is expected to be higher than the anxiety level after counseling. In symbols, $d = B - A$ should tend to be positive, and the population mean of differences μ_d also should be positive. Therefore, we have

$$H_0\colon \mu_d = 0 \quad \text{and} \quad H_1\colon \mu_d > 0$$

(b) Find the sample test statistic $\bar{d}$ and convert it to a corresponding test statistic t.
First we need to compute $\bar{d}$ and s_d. Using formulas or a calculator and the d values shown in Table 10-2, we find that

$$\bar{d} \approx 33.33 \quad \text{and} \quad s_d \approx 22.92$$

Using these values together with $n = 9$ and $\mu_d = 0$, we have

$$t = \frac{\bar{d} - 0}{(s_d/\sqrt{n})} \approx \frac{33.33}{22.92/\sqrt{9}} \approx 4.363$$

(c) Find the P-value for the test statistic and sketch the P-value on the t distribution.
Since we have a right-tailed test, the P-value is the area to the right of $t = 4.363$, as shown in Figure 10-1. In Table 4 of the Appendix, we find an interval containing the P-value. Use entries from the row headed by $d.f. = n - 1 = 9 - 1 = 8$. The test statistic $t = 4.636$ falls between 3.355 and 5.041. The P-value for the sample t falls between the corresponding one-tail areas 0.005 and 0.0005. (See Table 10-3, Excerpt from Table 4 of the Appendix.)

$$0.0005 < P\text{-value} < 0.005$$

FIGURE 10-1

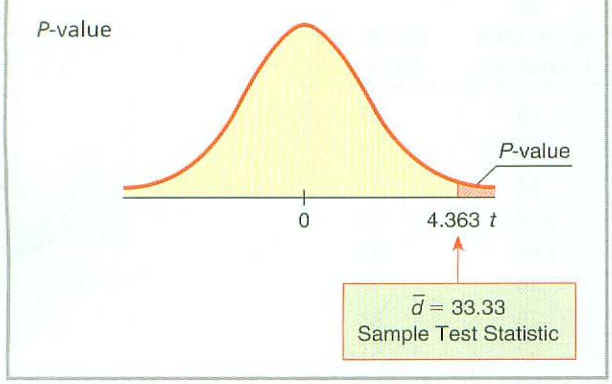


TABLE 10-3 Excerpt from Student's t Distribution Table (Table 4, Appendix)

✓ one-tail area	0.005	0.0005
two-tail area	0.010	0.0010
$d.f. = 8$	3.355	5.041
	↑ Sample t = 4.363	

(d) Conclude the test.

Since the interval containing the P-value lies to the left of $\alpha = 0.01$, we reject H_0. *Note:* Using the raw data and software, the P-value ≈ 0.0012.

(e) Interpret the results.
At the 1% level of significance, we conclude that the counseling sessions reduce the average anxiety level of patients about to undergo corrective heart surgery. ◆

The problem we have just solved is a paired difference problem of the "before and after" type. The next guided exercise demonstrates a paired difference problem of the "matched pair" type.

GUIDED EXERCISE 2

Paired difference test

Do educational toys make a difference in the age at which a child learns to read? To study this question, researchers designed an experiment in which one group of preschool children spent 2 hours each day (for 6 months) in a room well supplied with "educational" toys such as alphabet blocks, puzzles, ABC readers, coloring books featuring letters, and so forth. A control group of children spent 2 hours a day for 6 months in a "noneducational" toy room. It was anticipated that IQ differences and home environment might be uncontrollable factors unless identical twins could be used. Therefore, six pairs of identical twins of preschool age were randomly selected. From each pair, one member was randomly selected to participate in the experimental (i.e., educational toy room) group and the other in the control (i.e., noneducational toy room) group. For each twin the data item recorded is the age in months at which the child began reading at the primary level (Table 10-4).

TABLE 10-4 Reading Ages for Identical Twins in Months

Twin Pair	Experimental Group B = Reading Age	Control Group A = Reading Age	Difference $d = B - A$
1	58	60	
2	61	64	
3	53	52	
4	60	65	
5	71	75	
6	62	63	

(a) Compute the entries in the $d = B - A$ column of Table 10-4. Using formulas for the mean and sample standard deviation or a calculator with mean and sample standard deviation keys, compute $\bar{d}$ and s_d.

Pair	$d = B - A$
1	−2
2	−3
3	1
4	−5
5	−4
6	−1

$\bar{d} \approx -2.33$

$s_d \approx 2.16$

Continued

GUIDED EXERCISE 2 continued

(b) What is the null hypothesis?

H_0: $\mu_d = 0$

(c) To test the claim that the experimental group learned to read at a *different age* (either younger or older), what should the alternate hypothesis be?

H_1: $\mu_d \neq 0$

(d) Convert the sample test statistic $\bar{d}$ to a t value. Find the degrees of freedom.

Using $\mu_d = 0$ from H_0, $\bar{d} = -2.33$, $n = 6$, and $s_d = 2.16$, we get

$$t = \frac{\bar{d} - \mu_d}{(s_d/\sqrt{n})} \approx \frac{-2.33 - 0}{(2.16/\sqrt{6})} \approx -2.642$$

$$d.f. = n - 1 = 6 - 1 = 5$$

(e) When we use Table 4 of the Appendix to find an interval containing the P-value, do we use one-tail or two-tail areas? Why? Sketch a figure showing the P-value. Find an interval containing the P-value.

This is a two-tailed test, so we use two-tail areas.

FIGURE 10-2 P-value

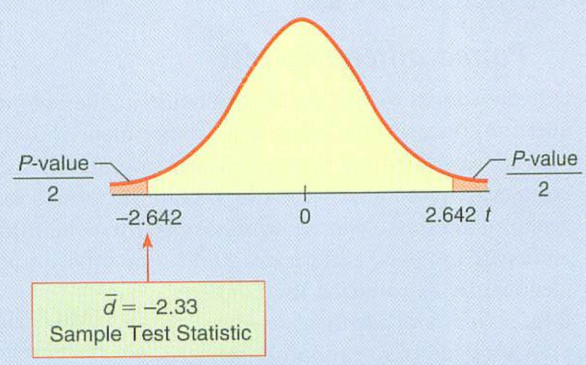


TABLE 10-5 Excerpt from Student's t Table

one-tail area	0.025	0.010
✓ two-tail area	0.050	0.020
$d.f.$ = 5	2.571	3.365

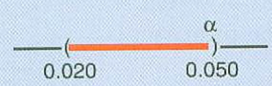
↑ Sample t = 2.642

The sample t is between 2.571 and 3.365.

$0.020 < P\text{-value} < 0.050$

(f) Using $\alpha = 0.05$, do we reject or fail to reject H_0? Interpret your results in the context of this application.

Since the interval containing the P-value has values that are all smaller than 0.05, we reject H_0.



At the 5% level of significance, the experiment indicates that educational toys make a difference in the age at which a child learns to read.

Note: Using the raw data and software, P-value ≈ 0.0457.

TECH NOTE Both Excel and Minitab support paired difference tests directly. On the TI-84Plus and TI-83Plus calculators, construct a column of differences and then do a t test on the data in that column. For each technology, be sure to relate the alternate hypothesis to the "before and after" assignments. All the displays show the results for the data of Guided Exercise 2.

TI-84Plus/TI-83Plus Enter the "before" data in column L1 and the "after" data in column L2. Highlight L3, type L1 − L2, and press Enter. The column L3 now contains the $B - A$ differences. To conduct the test, press **STAT**, select **TESTS**, and use option **2:T-Test**. Note that the letter x is used in place of d.

L1	L2	L3 3
58	60	-2
61	64	-3
53	52	1
60	65	-5
71	75	-4
62	63	-1

L3(7) =

```
T-Test
 μ≠0
 t=-2.645751311
 p=.0456591238
 x̄=-2.333333333
 Sx=2.160246899
 n=6
```

Excel Enter the data in two columns. Use the menu choices **Tools ➤ Data Analysis ➤ t-Test: Paired Two-Sample for Means.** Fill in the dialogue box with the hypothesized mean difference of 0. Set alpha.

	B	C	D
	t-Test: Paired Two Sample for Means		
		Variable 1	*Variable 2*
	Mean	60.83333	63.16666667
	Variance	34.96667	55.76666667
	Observations	6	6
	Pearson Correlation	0.974519	
	Hypothesized Mean Difference	0	
	df	5	
	t Stat	−2.64575	
	P(T<=t) one-tail	0.02283	
	t Critical one-tail	2.015049	
P-value →	P(T<=t) two-tail	0.045659	
	t Critical two-tail	2.570578	

Minitab Enter the data in two columns. Use the menu selection **Stat ➤ Basic Statistics ➤ Paired *t*.** Under Options, set the null and alternate hypotheses.

```
Paired T-Test and Confidence Interval
Paired T for B - A
              N      Mean      St Dev SE     Mean
B             6      60.83     5.91          2.41
A             6      63.17     7.47          3.05
Difference    6     -2.333     2.160         0.882
95% CI for mean difference: (-4.601, -0.066)
T-Test of mean difference = 0 (vs not = 0):
     T-Value = -2.65 P-Value = 0.046
```

Traditional Method Using Critical Regions (Optional)

For a fixed preset level of significance α, the P-value method of testing is logically equivalent to the critical region method of testing. This book emphasizes the P-value method because of its great popularity and because it is readily compatible with most computer software. However, for completeness, we provide an optional example utilizing the critical region method.

Recall that the critical region method and the P-value method of testing share a number of steps. Both methods use the same *null and alternate hypotheses,* the same *sample test statistic,* and the same *sampling distribution* for the test statistic. Instead of computing the P-value and comparing it to the level of significance α, the critical region method compares the sample test statistic to a *critical value* from the sampling distribution that is based on α and the alternate hypothesis (left-tailed, right-tailed, or two-tailed).

Test conclusions based on critical values

For a *right-tailed test,* if the sample *test statistic* $\geq$ *critical value,* reject H_0.

For a *left-tailed test,* if the sample *test statistic* $\leq$ *critical value,* reject H_0.

For a *two-tailed test,* if the sample *test statistic lies beyond the critical values* (that is, $\leq$ negative critical value or $\geq$ positive critical value), reject H_0.

Otherwise, in each case, do not reject H_0.

EXAMPLE 3
Critical region method

Let's revisit Guided Exercise 2 regarding educational toys and reading age and conclude the test using the critical region method. Recall that there were six pairs of twins. One twin of each set was given educational toys and the other was not. The difference d in reading ages for each pair of twins was measured, and $\alpha = 0.05$.

SOLUTION: From Guided Exercise 2, we have

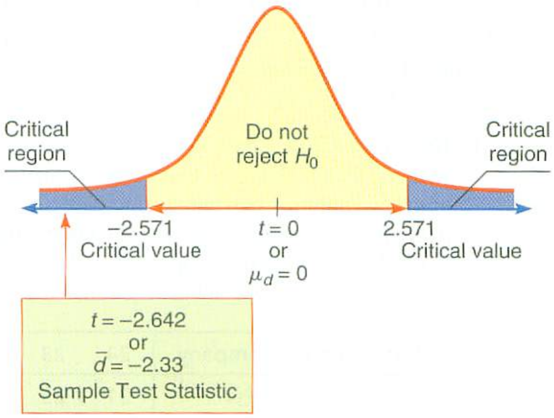
$$H_0\text{: } \mu_d = 0 \qquad \text{and} \qquad H_1\text{: } \mu_d \neq 0$$

We computed the sample test statistic $\bar{d} \approx -2.33$ with corresponding $t \approx -2.642$.

(a) Find the critical values for $\alpha = 0.05$.
Since the number of pairs is $n = 6$, $d.f. = n - 1 = 5$. In the Student's t distribution table (Table 4 of the Appendix), look in the row headed by 5. To find the column, locate $\alpha = 0.05$ in the *two-tail area* row, since we have a two-tailed test. The critical values are $\pm t_0 = \pm 2.571$.

FIGURE 10-3

Critical Region with $\alpha = 0.05$, $d.f. = 5$



(b) Sketch the critical regions and place the t value of the sample test statistic $\bar{d}$ on the sketch. Conclude the test. Compare the result to the result given by the P-value method of Guided Exercise 2.

Since the sample test statistic falls in the critical region (see Figure 10-3), we reject H_0 at the 5% level of significance. At this level, educational toys seem to make a difference in reading age. Notice that this conclusion is consistent with the conclusion obtained using the P-value. ◆

VIEWPOINT

DUI

DUI usually means "driving under the influence" of *alcohol,* but driving under the influence of *sleep loss* can be just as dangerous. Researchers in Australia have found that after staying awake for 24 hours straight, a person will be about as impaired as if he or she had had enough alcohol to be legally drunk in most U.S. states (Source: *Rocky Mountain News*). Using driver simulation exams and statistical tests (paired difference tests) found in this section, it is possible to show that the null hypothesis H_0: $\mu_d = 0$ cannot be rejected. Or put another way, the average level of impairment for a given individual from alcohol (at the DUI level) is about the same as the average level of impairment from sleep loss (24 hours without sleep).

SECTION 10.1 PROBLEMS

Please provide the following information for Problems 1–10.

(a) What is the level of significance? State the null and alternate hypotheses. Will you use a left-tailed, right-tailed, or two-tailed test?

(b) What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?

(c) Find (or estimate) the P-value. Sketch the sampling distribution and show the area corresponding to the P-value.

(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?

(e) State your conclusion in the context of the application.

In these problems, assume that the distribution of differences is approximately normal.

Note: For degrees of freedom *d.f.* not in the Student's t table, use the closest *d.f.* that is *smaller*. In some situations, this choice of *d.f.* may increase the P-value a small amount, and therefore produce a slightly more "conservative" answer.

1. *Business: CEO Raises* Are America's top chief executive officers (CEOs) really worth all that money? One way to answer this question is to look at row B, the annual company percentage increase in revenue, versus row A, the CEO's annual percentage salary increase at that same company (Source: *Forbes*, Vol. 159, No. 10). A random sample of companies such as John Deere & Co., General Electric, Union Carbide, and Dow Chemical yielded the following data:

B: Percent for company	24	23	25	18	6	4	21	37
A: Percent for CEO	21	25	20	14	−4	19	15	30

Do these data indicate that the population mean percentage increase in corporate revenue (row B) is different from the population mean percentage increase in CEO salary? Use a 5% level of significance.

2. *Fishing: Shore or Boat?* Is fishing better from a boat or from the shore? Pyramid Lake is on the Paiute Indian Reservation in Nevada. Presidents, movie stars, and people who just want to catch fish go to Pyramid Lake for really large cutthroat trout. Let row B represent hours to catch a fish fishing from the shore, and let row A represent hours to catch a fish using a boat. The following data are paired by month from October through April (Source: *Pyramid Lake Fisheries*, Paiute Reservation, Nevada).

	Oct.	Nov.	Dec.	Jan.	Feb.	March	April
B: Shore	1.6	1.8	2.0	3.2	3.9	3.6	3.3
A: Boat	1.5	1.4	1.6	2.2	3.3	3.0	3.8

Use a 1% level of significance to test if there is a difference in the population mean hours to catch a fish using a boat compared with fishing from the shore.

3. *Ecology: Rocky Mountain National Park* The following is based on information taken from *Winter Wind Studies in Rocky Mountain National Park*, by D. E. Glidden (Rocky Mountain Nature Association). At five weather stations on Trail Ridge Road in Rocky Mountain National Park, the peak wind gusts (in miles per hour) for January and April are recorded below.

Weather Station	1	2	3	4	5
January	139	122	126	64	78
April	104	113	100	88	61

Does this information indicate that the peak wind gusts are higher in January than in April? Use $\alpha = 0.01$.

4. *Wildlife: Highways* The western United States has a number of four-lane interstate highways that cut through long tracts of wilderness. To prevent car accidents with wild animals, the highways are bordered on both sides with 12-foot-high woven wire fences.

Although the fences prevent accidents, they also disturb the winter migration pattern of many animals. To compensate for this disturbance, the highways have frequent wilderness underpasses designed for exclusive use by deer, elk, and other animals.

In Colorado, there is a large group of deer that spend their summer months in a region on one side of a highway and survive the winter months in a lower region on the other side. To determine if the highway has disturbed deer migration to the winter feeding area, the following data were gathered from a random sample of 10 wilderness districts in the winter feeding area. Row *B* represents the average January deer count for a 5-year period before the highway was built, and row *A* represents the average January deer count for a 5-year period after the highway was built. The highway department claims that the January population has not changed. Test this against the claim that the January population has dropped. Use a 5% level of significance. Units used in the table are hundreds of deer.

Wilderness District	1	2	3	4	5	6	7	8	9	10
B: Before highway	10.3	7.2	12.9	5.8	17.4	9.9	20.5	16.2	18.9	11.6
A: After highway	9.1	8.4	10.0	4.1	4.0	7.1	15.2	8.3	12.2	7.3

5. *Wildlife: Wolves* In environmental studies, sex ratios are of great importance. Wolf society, packs, and ecology have been studied extensively at different locations in the U.S. and foreign countries. Sex ratios for eight study sites in northern Europe are shown below (based on *The Wolf* by L. D. Mech, University of Minnesota Press).

Gender Study of Large Wolf Packs

Location of Wolf Pack	% Males (Winter)	% Males (Summer)
Finland	72	53
Finland	47	51
Finland	89	72
Lapland	55	48
Lapland	64	55
Russia	50	50
Russia	41	50
Russia	55	45

It is hypothesized that in winter, "loner" males (males not present in summer packs) join the pack to increase survival rate. Use a 5% level of significance to test the claim that the average percentage of males in a wolf pack is higher in winter.

6. *Demographics: Birth Rate and Death Rate* In the following data pairs, *A* represents birth rate and *B* represents death rate per 1000 resident population. The data are paired by counties in the Midwest. A random sample of 16 counties gave the following information (Reference: *County and City Data Book,* U.S. Department of Commerce).

A:	12.7	13.4	12.8	12.1	11.6	11.1	14.2	15.1
B:	9.8	14.5	10.7	14.2	13.0	12.9	10.9	10.0

A:	12.5	12.3	13.1	15.8	10.3	12.7	11.1	15.7
B:	14.1	13.6	9.1	10.2	17.9	11.8	7.0	9.2

Do the data indicate a difference (either way) between population average birth rate and death rate in this region? Use $\alpha = 0.01$.

7. *Golf: Tournaments* Do professional golfers play better in their first round? Let row *B* represent the score in the fourth (and final) round, and let row *A* represent the score in the first round of a professional golf tournament. A random sample of finalists in the British Open gave the following data for first and last rounds in the tournament (Source: *Golf Almanac*).

B: Last	73	68	73	71	71	72	68	68	74
A: First	66	70	64	71	65	71	71	71	71

Do the data indicate that the population mean score on the last round is higher than that on the first? Use a 5% level of significance.

8. *Archaeology: Stone Tools* The following is based on information taken from *Bandelier Archaeological Excavation Project: Summer 1990 Excavations at Burnt Mesa Pueblo and Casa del Rito*, edited by T. A. Kohler (Washington State University, Department of Anthropology). The artifact frequency for an excavation of a kiva in Bandelier National Monument gave the following information.

Stratum	Flaked Stone Tools	Nonflaked Stone Tools
1	7	3
2	3	2
3	10	1
4	1	3
5	4	7
6	38	32
7	51	30
8	25	12

Does this information indicate that there tend to be more flaked stone tools than nonflaked stone tools at this excavation site? Use a 5% level of significance.

9. *Archaeology: Pot Sherds* From the same stratum of an excavated block of rooms at Bandelier National Monument, the following information was obtained for number of sherds of service ware found in two different subareas (see reference in Problem 8):

Service Ware	Subarea 1	Subarea 2
Socorro black-on-white	10	4
Santa Fe black-on-white	42	39
Galisteo black-on-white	15	21
Puerco black-on-red	6	9
Wingate black-on-red	11	6

Does this information indicate that there is a difference (either higher or lower) in the population mean number of service ware sherds in subarea 1 compared with subarea 2? Use a 5% level of significance.

10. *Economics: Cost of Living Index* In the following data pairs, A represents the cost of living index for housing and B represents the cost of living index for groceries. The data are paired by metropolitan areas of the United States. A random sample of 36 metropolitan areas gave the following information (Reference: *Statistical Abstract of the United States*, 121st edition).

A:	132	109	128	122	100	96	100	131	97
B:	125	118	139	104	103	107	109	117	105

A:	120	115	98	111	93	97	111	110	92
B:	110	109	105	109	104	102	100	106	103

A:	85	109	123	115	107	96	108	104	128
B:	98	102	100	95	93	98	93	90	108

A:	121	85	91	115	114	86	115	90	113
B:	102	96	92	108	117	109	107	100	95

i. Let d be the random variable $d = A - B$. Use a calculator to verify that $\bar{d} \approx 2.472$ and $s_d \approx 12.124$.

ii. Do the data indicate that the U.S. population mean cost of living index for housing is higher than that for groceries in these areas? Use $\alpha = 0.05$.

11. *Critical Region Method: Student's t* Solve Problem 1 using the critical region method of testing. Compare your conclusions with the conclusions obtained by using the P-value method. Are they the same?

12. *Critical Region Method: Student's t* Solve Problem 3 using the critical region method of testing. Compare your conclusions with the conclusions obtained by using the P-value method. Are they the same?

10.2 Inferences About the Difference of Two Means $\mu_1 - \mu_2$

FOCUS POINTS

- Identify independent samples and sampling distributions.
- Compute the sample test statistic and P-value for testing $\mu_1 - \mu_2$.
- Find confidence intervals for $\mu_1 - \mu_2$.

Independent Samples

Many practical applications of statistics involve a comparison of two population means or two population proportions. In Section 10.1, we considered tests of difference of means for *dependent samples*. With dependent samples, we could pair the data and then consider the differences of the data measurements d. In this section, we will turn our attention to inferences regarding differences of means from *independent samples.*

Two samples are **independent** if the selection of sample data from one population is completely unrelated to the selection of sample data from the other population.

Independent samples occur very naturally when we draw *two random samples,* one from the first population and one from the second population. Because *both* samples are random samples, there is no pairing of measurements between the two populations.

GUIDED EXERCISE 3

Distinguish between independent and dependent samples

For each experiment, categorize the sampling as independent or dependent, and explain your choice.

(a) In many medical experiments, a sample of subjects is randomly divided into two groups. One group is given a specific treatment, and the other group is given a placebo. After a certain period of time, both groups are measured for the same condition. Do the measurements from these two groups constitute independent or dependent samples?

→ Since the subjects are *randomly assigned* to the two treatment groups (one receives a treatment, the other a placebo), the resulting measurements would form independent samples.

(b) In an accountability study, a group of students in an English composition course is given a pretest. After the course, the same students are given a posttest covering similar material. Are the two groups of scores independent or dependent?

→ Since the pretest scores and the posttest scores are from the same students, the samples are dependent. Each student has both a pretest score and a posttest score, so there is a natural pairing of data values.

Hypothesis Tests and Confidence Intervals for $\mu_1 - \mu_2$ (σ_1 and σ_2 known)

The $\bar{x}_1 - \bar{x}_2$ sampling distribution

In this section, we will use probability distributions that arise from a difference of means. How do we obtain such distributions? Suppose that we have two statistical variables x_1 and x_2, each with its own distribution. We take *independent* random samples of size n_1 from the x_1 distribution and of size n_2 from the x_2 distribution. Then we compute the respective means $\bar{x}_1$ and $\bar{x}_2$. Now consider the difference $\bar{x}_1 - \bar{x}_2$. This expression represents a difference of means. If we repeat this sampling process over and over, we will create lots of $\bar{x}_1 - \bar{x}_2$ values. Figure 10-4 illustrates the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

The values of $\bar{x}_1 - \bar{x}_2$ that come from repeated (independent) sampling of populations 1 and 2 can be arranged in a relative-frequency table and a relative-frequency histogram (see Section 2.1). This would give us an experimental idea of the theoretical probability distribution of $\bar{x}_1 - \bar{x}_2$.

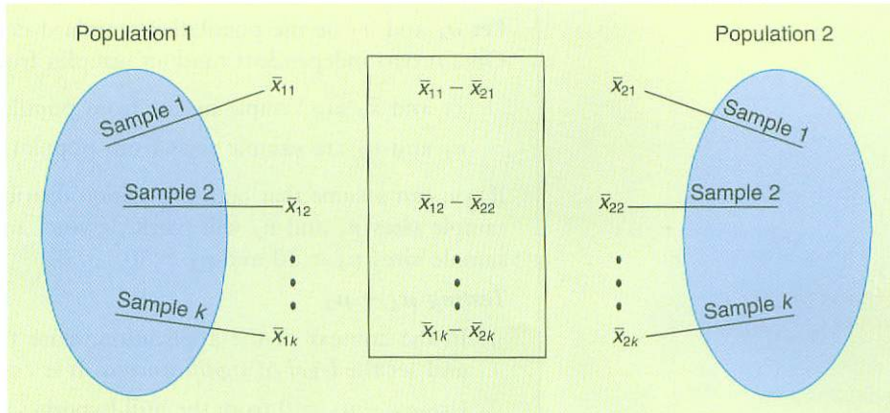
Fortunately, it is not necessary to carry out this lengthy process for each example. The results have been worked out mathematically. The next theorem presents the main results.

◇ **THEOREM 10.2** Let x_1 and x_2 have normal distributions with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively. If we take independent random samples of size n_1 from the x_1 distribution and of size n_2 from the x_2 distribution, then the variable $\bar{x}_1 - \bar{x}_2$ has

1. a normal distribution
2. mean $\mu_1 - \mu_2$
3. standard deviation $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ ◇

FIGURE 10-4

Sampling Distribution of $\bar{x}_1 - \bar{x}_2$



◇ **COMMENT** The theorem requires that x_1 and x_2 have *normal* distributions. However, if *both* n_1 and n_2 are 30 or larger, then the central limit theorem (Section 7.5) assures us that $\bar{x}_1$ and $\bar{x}_2$ are approximately normally distributed. In this case, the conclusions of the theorem are again valid even if the original x_1 and x_2 distributions are not exactly normal. ◇

Theorem 10.2 gives us the basis for hypothesis tests and confidence intervals for $\mu_1 - \mu_2$ when σ_1 and σ_2 are both known.

Hypotheses for testing difference of means

When testing the difference of means, it is customary to use the null hypothesis

$$H_0: \mu_1 - \mu_2 = 0 \text{ or, equivalently, } H_0: \mu_1 = \mu_2$$

As mentioned in Section 9.1, the null hypothesis is set up to see if it can be rejected. When testing the difference of means, we first set up the hypothesis H_0 that there is no difference. The alternate hypothesis could then be any of the ones listed in Table 10-6. The alternate hypothesis and consequent type of test used depend on the particular problem. Note that μ_1 is always listed first.

Using Theorem 10.2 and the central limit theorem (Section 7.5), we can summarize the procedure for testing $\mu_1 - \mu_2$ and finding a confidence interval when both σ_1 and σ_2 are known.

TABLE 10-6 Alternate Hypotheses and Type of Test: Difference of Two Means

	H_1		Type of Test
$H_1: \mu_1 - \mu_2 < 0$	or equivalently	$H_1: \mu_1 < \mu_2$	Left-tailed test
$H_1: \mu_1 - \mu_2 > 0$	or equivalently	$H_1: \mu_1 > \mu_2$	Right-tailed test
$H_1: \mu_1 - \mu_2 \neq 0$	or equivalently	$H_1: \mu_1 \neq \mu_2$	Two-tailed test

PROCEDURE

How to test $\mu_1 - \mu_2$ and find a confidence interval when both σ_1 and σ_2 are known

Let σ_1 and σ_2 be the population standard deviations of populations 1 and 2. Obtain two independent random samples from populations 1 and 2, where

$\bar{x}_1$ and $\bar{x}_2$ are sample means from populations 1 and 2

n_1 and n_2 are sample sizes from populations 1 and 2

If you can assume that both population distributions 1 and 2 are normal, any sample sizes n_1 and n_2 will work. If you cannot assume this, then use sample sizes $n_1 \geq 30$ and $n_2 \geq 30$.

Hypothesis test for $\mu_1 - \mu_2$ (σ_1 and σ_2 known)

Testing $\mu_1 - \mu_2$

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α. It is customary to use H_0: $\mu_1 - \mu_2 = 0$.
2. Use $\mu_1 - \mu_2 = 0$ from the null hypothesis together with $\bar{x}_1$, $\bar{x}_2$, σ_1, σ_2, n_1, and n_2 to compute the sample *test statistic*.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

3. Use the standard normal distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample 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. *State your conclusion* in the context of the application.

Confidence interval for $\mu_1 - \mu_2$ (σ_1 and σ_2 known)

Confidence interval for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

where $E = z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

c = confidence level ($0 < c < 1$)

z_c = critical value for confidence level c based on the standard normal distribution (See Table 3(b) of the Appendix for commonly used values.)

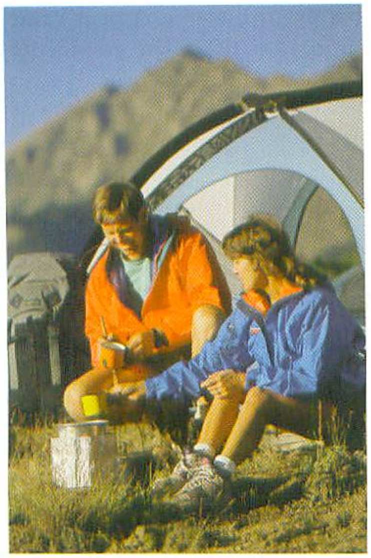
EXAMPLE 4

Testing the difference of means and finding a confidence interval (σ_1 and σ_2 known)

A consumer group is testing camp stoves. To test the heating capacity of a stove, it measures the time required to bring 2 quarts of water from 50°F to boiling (at sea level). Two competing models are under consideration. Ten stoves of the first model and 12 stoves of the second model are tested. The following results are obtained.

Model 1: Mean time $\bar{x}_1 = 11.4$ min; $\sigma_1 = 2.5$ min; $n_1 = 10$

Model 2: Mean time $\bar{x}_2 = 9.9$ min; $\sigma_2 = 3.0$ min; $n_2 = 12$



Assume that the time required to bring water to a boil is normally distributed for each stove.

Hypothesis test: Is there any difference (either way) between the performances of these two models? Use a 5% level of significance.

SOLUTION:

(a) State the null and alternate hypotheses and note the value of α.
Let μ_1 and μ_2 be the means of the distributions of times for models 1 and 2, respectively. We set up the null hypothesis to state that there is no difference:

$$H_0: \mu_1 = \mu_2 \quad \text{or} \quad H_0: \mu_1 - \mu_2 = 0$$

The alternate hypothesis states that there is a difference:

$$H_1: \mu_1 \neq \mu_2 \quad \text{or} \quad H_1: \mu_1 - \mu_2 \neq 0$$

The level of significance is $\alpha = 0.05$.

(b) Compute the sample test statistic $\bar{x}_1 - \bar{x}_2$ and then convert it to a z value.
Note that we use the standard normal distribution because the original distributions are normal and the standard deviations are known.

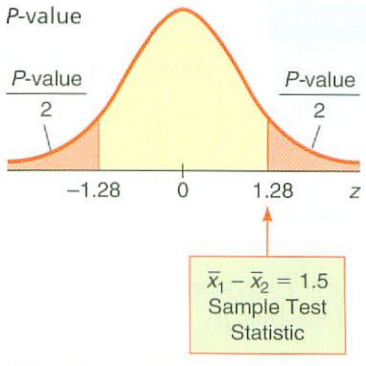
We are given the values $\bar{x}_1 = 11.4$ and $\bar{x}_2 = 9.9$. Therefore, the sample test statistic is $\bar{x}_1 - \bar{x}_2 = 11.4 - 9.9 = 1.5$. To convert this to a z value, we use the values $\sigma_1 = 2.5$, $\sigma_2 = 3.0$, $n_1 = 10$, and $n_2 = 12$. From the null hypothesis, $\mu_1 - \mu_2 = 0$.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1.5}{\sqrt{\frac{2.5^2}{10} + \frac{3.0^2}{12}}} \approx 1.28$$

(c) Find the P-value and sketch the area on the standard normal curve.
Figure 10-5 shows the P-value. Use the standard normal distribution (Table 3 of the Appendix) and the fact that we have a two-tailed test. P-value $\approx 2(0.1003) = 0.2006$.

FIGURE 10-5

P-value



(d) Conclude the test.
The P-value is 0.2006 and $\alpha = 0.05$. Since P-value $> \alpha$, do not reject H_0.

(e) Interpret the results.
At the 5% level of significance, the sample data do not indicate any difference in the population mean times for boiling water for the two stove models.

Confidence interval: Find a 95% confidence interval for the population difference $\mu_1 - \mu_2$ of mean times to boil water for the two stoves. Interpret the results.

SOLUTION:

Note that both σ_1 and σ_2 are known and that the times for each stove to bring 2 quarts of water to a boil are normally distributed. Therefore, by Theorem

10.2, $\bar{x}_1 - \bar{x}_2$ is normal. We use the normal distribution to find the critical value $z_{0.95}$. From Table 3(b) of the Appendix, we see that $z_{0.95} = 1.96$. The confidence interval is

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

where $\bar{x}_1 - \bar{x}_2 = 11.4 - 9.9 = 1.5$ min and

$$E = z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.96\sqrt{\frac{2.5^2}{10} + \frac{3.0^2}{12}} \approx 2.30 \text{ min}$$

The confidence interval is

$$-0.8 \text{ min} < \mu_1 - \mu_2 < 3.8 \text{ min}$$

Notice that the confidence interval contains both negative and positive values. At the 95% confidence level, we cannot conclude that μ_1 is either less than or greater than μ_2. This result is consistent with the conclusion of the hypothesis test—that is, at the 5% level, the evidence is not sufficient to conclude that there is a difference in the average times for boiling 2 quarts of water between the two stoves. ◇

Meaning of confidence interval for $\mu_1 - \mu_2$

In Example 4 we saw that the confidence interval for $\mu_1 - \mu_2$ contained *both* positive and negative numbers. At the 95% level of confidence, we could not conclude that there was a difference in value between μ_1 and μ_2. There are two other cases: either all the values in the confidence interval are positive, or they are all negative. The next procedure summarizes the interpretation of all three cases.

PROCEDURE

How to interpret confidence intervals for differences

Suppose that we construct a $c\%$ confidence interval for $\mu_1 - \mu_2$. Then three cases arise:

1. The $c\%$ confidence interval contains only *negative values*. In this case, we conclude that $\mu_1 - \mu_2 < 0$, and we are therefore $c\%$ confident that $\mu_1 < \mu_2$.
2. The $c\%$ confidence interval contains only *positive values*. In this case, we conclude that $\mu_1 - \mu_2 > 0$, and we can be $c\%$ confident that $\mu_1 > \mu_2$.
3. The $c\%$ confidence interval contains *both positive and negative values*. In this case, we cannot at the $c\%$ confidence level conclude that either μ_1 or μ_2 is larger. However, if we *reduce* the confidence level c to a *smaller value*, then the confidence interval will, in general, be shorter (explain why). A shorter confidence interval *might* put us back into case 1 or case 2 above (again, explain why).

GUIDED EXERCISE 4

Interpret a confidence interval

(a) A study reported a 90% confidence interval for the difference of means to be

$10 < \mu_1 - \mu_2 < 20$

For this interval, what can you conclude about the respective values of μ_1 and μ_2?

➡ At a 90% level of confidence, we can say that the difference $\mu_1 - \mu_2$ is positive, so $\mu_1 - \mu_2 > 0$ and $\mu_1 > \mu_2$.

(b) A study reported a 95% confidence interval for the difference of means to be

$-0.32 < \mu_1 - \mu_2 < 0.16$

From this interval, what can you conclude about the respective values of μ_1 and μ_2?

➡ At the 95% confidence level, we see that the difference of means ranges from negative to positive values. We cannot tell from this interval if μ_1 is greater than μ_2 or μ_1 is less than μ_2.

◇ **COMMENT:** In the case of large samples ($n_1 \geq 30$ and $n_2 \geq 30$), it is not unusual to see σ_1 and σ_2 approximated by s_1 and s_2. Then Theorem 10.2 is used as a basis for approximating confidence intervals for $\mu_1 - \mu_2$. In other words, when samples are large, sample estimates for σ_1 and σ_2 can be used together with the standard normal distribution to test $\mu_1 - \mu_2$ and find confidence intervals. However, in this text, we follow the more common convention of using a Student's t distribution whenever σ_1 and σ_2 are unknown. ◇

Hypothesis Tests and Confidence Intervals for $\mu_1 - \mu_2$ When σ_1 and σ_2 Are Unknown

When σ_1 and σ_2 are unknown, we turn to a Student's t distribution. As before, when we use a Student's t distribution, we require that our populations be normal or approximately normal (mound-shaped and symmetric) when the sample sizes n_1 and n_2 are less than 30. We also replace σ_1 by s_1 and σ_2 by s_2. Then we consider the approximate t value, attributed to Welch (*Biometrika,* Vol. 29, pp. 350–362).

$$t \approx \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Unfortunately, this approximation is *not* exactly a Student's t distribution. However, it will be a good approximation provided we adjust the degrees of freedom by one of the following methods.

1. The adjustment for the degrees of freedom is calculated from sample data. The formula is called *Satterthwaite's approximation.* It is rather complicated. Satterthwaite's approximation is used in statistical software packages such as Minitab and in the TI-84Plus/TI-83Plus calculators. See Problem 13 for the formula.

2. An alternative method, which is much simpler, is to approximate the degrees of freedom using the *smaller* of $n_1 - 1$ and $n_2 - 1$.

For hypothesis tests and confidence intervals, we take the degrees of freedom *d.f.* to be the smaller of $n_1 - 1$ and $n_2 - 1$. This commonly used choice for the degrees of freedom is more conservative that Satterthwaite's approximation in the sense that it produces a slightly larger *P*-value for testing or a slightly larger margin of error for confidence intervals.

Applying methods similar to those used for hypothesis tests and confidence intervals for μ when σ is unknown, and using the Welch approximation for *t*, we obtain the following results.

PROCEDURE

How to test $\mu_1 - \mu_2$ and find a confidence interval when σ_1 and σ_2 are unknown

Obtain two independent random samples from populations 1 and 2, where

$\bar{x}_1$ and $\bar{x}_2$ are sample means from populations 1 and 2

s_1 and s_2 are sample standard deviations from populations 1 and 2

n_1 and n_2 are sample sizes from populations 1 and 2

If you can assume that both population distributions 1 and 2 are normal or at least mound-shaped and symmetric, then any sample sizes n_1 and n_2 will work. If you cannot assume this, then use sample sizes $n_1 \geq 30$ and $n_2 \geq 30$.

Hypothesis test for $\mu_1 - \mu_2$ (σ_1 and σ_2 unknown)

Testing $\mu_1 - \mu_2$

1. In the context of the application, state the *null and alternate hypotheses* and set the *level of significance* α. It is customary to use H_0: $\mu_1 - \mu_2 = 0$.
2. Use $\mu_1 - \mu_2 = 0$ from the null hypothesis together with $\bar{x}_1$, $\bar{x}_2$, s_1, s_2, n_1, and n_2 to compute the sample *test statistic.*

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

   The sample test statistic distribution is approximately that of a Student's *t* with *degrees of freedom d.f.* = *smaller* of $n_1 - 1$ and $n_2 - 1$. Note that statistical software gives a slightly more accurate and larger *d.f.* based on Satterthwaite's approximation (see Problem 13).

3. Use a Student's *t* distribution and the type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample 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. *State your conclusion* in the context of the application.

Confidence interval for $\mu_1 - \mu_2$ (σ_1 and σ_2 unknown)

Confidence interval for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

where $E \approx t_c\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

c = confidence level ($0 < c < 1$)

t_c = critical value for confidence level c (See Table 4 of the Appendix.)

d.f. = *smaller* of $n_1 - 1$ and $n_2 - 1$. Note that statistical software gives a slightly more accurate and larger *d.f.* based on Satterthwaite's approximation.

EXAMPLE 5

Testing the difference of means and finding a confidence interval (σ_1 and σ_2 unknown)

Two competing headache remedies claim to give fast-acting relief. An experiment was performed to compare the mean lengths of time required for bodily absorption of brand A and brand B headache remedies.

Twelve people were randomly selected and given an oral dosage of brand A. Another 12 were randomly selected and given an equal dosage of brand B. The lengths of time in minutes for the drugs to reach a specified level in the blood were recorded. The means, standard deviations, and sizes of the two samples follow.

Brand A: $\bar{x}_1 = 21.8$ min; $s_1 = 8.7$ min; $n_1 = 12$

Brand B: $\bar{x}_2 = 18.9$ min; $s_2 = 7.5$ min; $n_2 = 12$

Past experience with the drug composition of the two remedies permits researchers to assume that both distributions are approximately normal.

Hypothesis test: Use a 5% level of significance to test the claim that there is no difference in the mean time required for bodily absorption. Also, find or estimate the *P*-value of the sample test statistic.

SOLUTION:

(a) $\alpha = 0.05$. The null hypothesis is

$$H_0: \mu_1 = \mu_2 \quad \text{or} \quad H_0: \mu_1 - \mu_2 = 0$$

Since we have no prior knowledge about which brand is faster, the alternate hypothesis is

$$H_1: \mu_1 \neq \mu_2 \quad \text{or} \quad H_0: \mu_1 - \mu_2 \neq 0$$

(b) Compute the sample test statistic.
We're given $\bar{x}_1 = 21.8$ and $\bar{x}_2 = 18.9$, so the sample difference is $\bar{x}_1 - \bar{x}_2 = 21.8 - 18.9 = 2.9$. Using $s_1 = 8.7$, $s_2 = 7.5$, $n_1 = 12$, $n_2 = 12$, and $\mu_1 - \mu_2 = 0$ from H_0, we compute the sample test statistic.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{2.9}{\sqrt{\dfrac{8.7^2}{12} + \dfrac{7.5^2}{12}}} \approx 0.875$$

(c) Estimate the *P*-value and sketch the area on a *t* graph.
Figure 10-6 shows the *P*-value. The degrees of freedom is $d.f. = 11$ (since both samples are of size 12). Because the test is a two-tailed test, the *P*-value is the area to the right of 0.875 together with the area to the left of -0.875. In the Student's *t* distribution table (Table 4 of the Appendix), we find an interval containing the *P*-value. Find 0.875 in the row headed by $d.f. = 11$. The test statistic 0.875 falls between the entries 0.697 and 1.214. Because this is a two-tailed test, we use the corresponding *P*-values 0.500 and 0.250 from the *two-tail area* row (see Table 10-7, Excerpt from Table 4). The *P*-value for the sample *t* is in the interval

$$0.250 < P\text{-value} < 0.500$$

(d) Conclude the test.

Since the interval containing the *P*-value lies to the right of $\alpha = 0.05$, we fail to reject H_0.

FIGURE 10-6

P-value

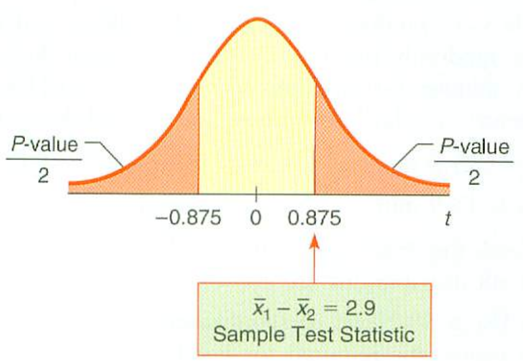


TABLE 10-7 Excerpt from Table 4 of the Appendix

one-tail area	0.250	0.125
✓two-tail area	0.500	0.250
d.f. = 11	0.697	1.214
	↑ Sample $t = 0.875$	

Note: Using Satterthwaite's approximation for the degrees of freedom, $d.f. \approx 21.53$, the *P*-value ≈ 0.3915. This value is in the interval we computed.

(e) Interpret the results.
At the 5% level of significance, there is insufficient evidence to conclude that there is a difference in mean times for the remedies to reach the specified level in the bloodstream.

Confidence interval: Find a 95% confidence interval for the population difference $\mu_1 - \mu_2$ of mean times for the competing drugs to reach a specified level in the blood. Interpret the results.

SOLUTION:

Because neither σ_1 nor σ_2 is known and because the distributions of times for each drug to enter the bloodstream are approximately normal, it is appropriate to use a Student's t distribution. We approximate the degrees of freedom by using the smaller of $n_1 - 1$ and $n_2 - 1$. Since n_1 and n_2 are both 12, $d.f. = 12 - 1 = 11$. From Table 4 of the Appendix, we find that $t_{0.95} = 2.201$. The confidence interval is

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

with $\bar{x}_1 - \bar{x}_2 = 21.8 - 18.9 = 2.9$ min and

$$E = t_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.201\sqrt{\frac{8.7^2}{12} + \frac{7.5^2}{12}} \approx 7.3 \text{ min}$$

The confidence interval is

$$-4.4 \text{ min} < \mu_1 - \mu_2 < 10.2 \text{ min}$$

Note: Using Satterthwaite's formula, $d.f. = 21.53$ and the 95% confidence interval is from −3.99 to 9.79 minutes.

Since the confidence interval contains both negative and positive values, we conclude that at the 95% confidence level, there is no difference in the population mean times for the two drugs to reach the bloodstream. This result is consistent with the two-tailed hypothesis test at the 5% level of significance. ◆

GUIDED EXERCISE 5

Testing the difference of means and finding a confidence interval (σ_1 and σ_2 unknown)

Suppose the experiment to measure the time in minutes for the headache remedies to enter the bloodstream (Example 5) yielded sample means, sample standard deviations, and sample sizes as follows:

Brand A: $\bar{x}_1 = 20.1$ min; $s_1 = 8.7$ min; $n_1 = 12$

Brand B: $\bar{x}_2 = 11.2$ min; $s_2 = 7.5$ min; $n_2 = 8$

Brand B claims to be faster.

Hypothesis test: Is this claim justified at the 5% level of significance? (Use the following steps to obtain the answer.)

(a) What is α? State H_0 and H_1.

→ $\alpha = 0.05$.

H_0: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 > \mu_2$ (or H_1: $\mu_1 - \mu_2 > 0$). This says that the mean time for brand B is less than the mean time for brand A.

(b) Compute the sample test statistic $\bar{x}_1 - \bar{x}_2$ and convert it to a t value.

→ $\bar{x}_1 - \bar{x}_2 = 20.1 - 11.2 = 8.9$. Using $s_1 = 8.7$, $s_2 = 7.5$, $n_1 = 12$, $n_2 = 8$, and $\mu_1 - \mu_2 = 0$ from H_0, we have

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{8.9}{\sqrt{\frac{8.7^2}{12} + \frac{7.5^2}{8}}} \approx 2.437$$

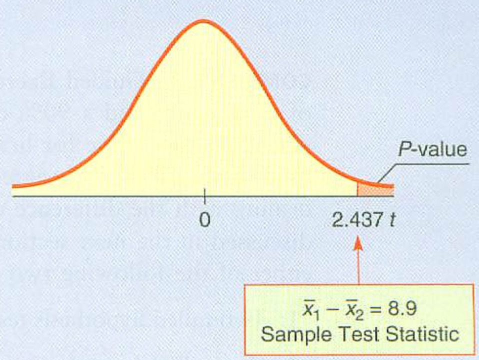
(c) What degrees of freedom do you use? To find an interval containing the P-value, do you use one-tail or two-tail areas of Table 4 of the Appendix? Sketch a figure showing the P-value. Find an interval for the P-value.

→ Since $n_2 < n_1$, $d.f. = n_2 - 1 = 8 - 1 = 7$. Use *one-tail area* of Table 4 of the Appendix.

TABLE 10-8 Excerpt from Student's t Table

✓one-tail area	0.025	0.010
two-tail area	0.050	0.020
d.f. = 7	2.365	2.998
	↑ Sample t = 2.437	

FIGURE 10-7 *P*-value



The sample t is between 2.365 and 2.998.

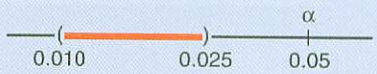
$0.010 < P\text{-value} < 0.025$

Continued

GUIDED EXERCISE 5 continued

(d) Do we reject or fail to reject H_0?

Since the interval containing the P-value has values that are all less than 0.05, we reject H_0.



Note: On the calculator with Satterthwaite's approximation for $d.f.$, we have $d.f. = 16.66$ and $P\text{-value} \approx 0.013$.

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

At the 5% level of significance, there is evidence that the mean time for brand B to enter the bloodstream is less than the mean time for brand A.

Confidence interval: Find a 90% confidence interval for the difference of population means $\mu_1 - \mu_2$ of times for the two competing headache remedies to enter the bloodstream.

(f) Find the degrees of freedom and the critical value $t_{0.90}$.

Since $n_2 < n_1$, $d.f. = n_2 - 1 = 8 - 1 = 7$.
From Table 4 of the Appendix, $t_{0.90} = 1.895$.

(g) Find the maximal error of estimate E for a 90% confidence interval.

$$E = t_c\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.895\sqrt{\frac{8.7^2}{12} + \frac{7.5^2}{8}} \approx 6.9 \text{ min}$$

(h) Find the sample difference $\bar{x}_1 - \bar{x}_2$ and a 90% confidence interval for $\mu_1 - \mu_2$.

$\bar{x}_1 - \bar{x}_2 = 20.1 - 11.2 = 8.9$ min

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$
$$8.9 - 6.9 < \mu_1 - \mu_2 < 8.9 + 6.9$$
$$2 \text{ min} < \mu_1 - \mu_2 < 15.8 \text{ min}$$

With Satterthwaite's formula, $d.f. = 16.66$ and the 90% confidence interval is 2.54 to 15.26 min.

(i) Interpret the results in the context of the application.

The confidence interval contains all positive numbers. This indicates that at the 90% confidence level, $\mu_1 > \mu_2$. It seems to take longer for brand A to enter the bloodstream.

◇ **COMMENT:** In Guided Exercise 5, we used a right-tailed test at the 5% level of significance and a 90% confidence interval. With both techniques, we concluded that the time for brand A to enter the bloodstream was longer than the time for brand B. This agreement of results is not an accident. In general, when dealing with the difference of means (and difference of proportions, as discussed in the next section), we will arrive at the same conclusion using either of the following two combinations.

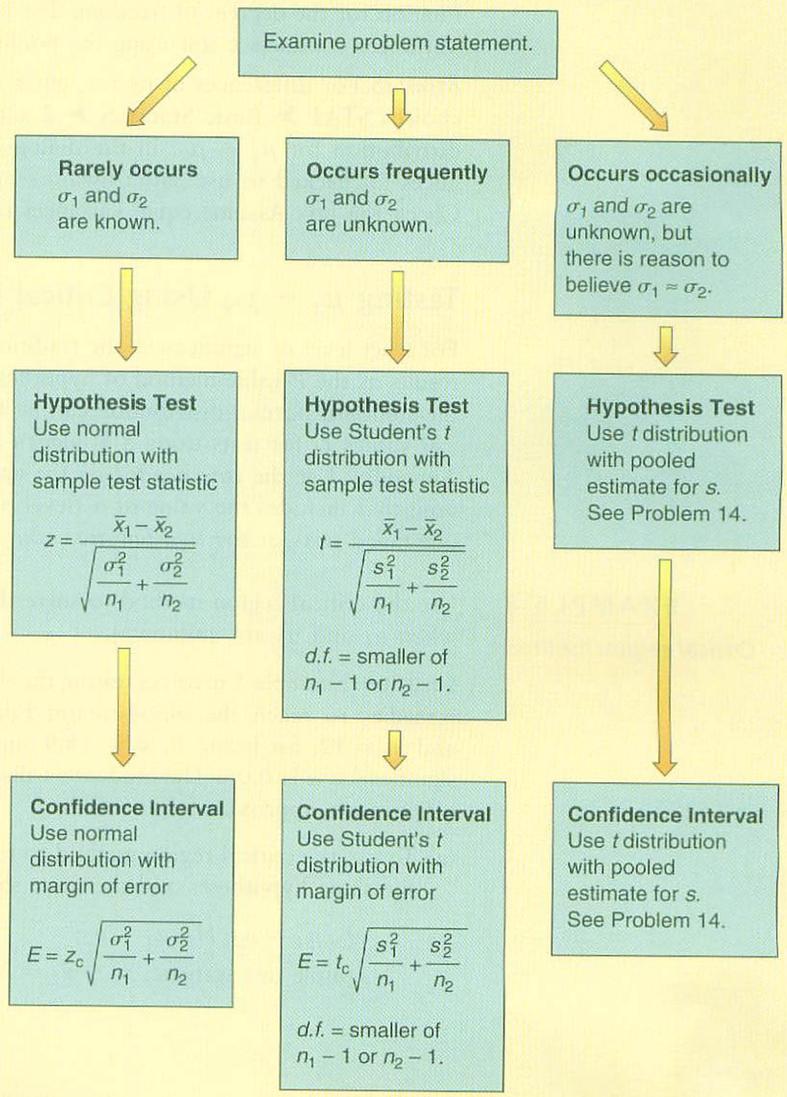
1. Two-tailed hypothesis test at level of significance α and $(1 - \alpha)$ confidence interval
2. One-tailed hypothesis test at level of significance α and $(1 - 2\alpha)$ confidence interval ◇

Alternate method using pooled standard deviation

There is another method for testing $\mu_1 - \mu_2$ and finding confidence intervals when σ_1 and σ_2 are unknown. Suppose that the sample values s_1 and s_2 are sufficiently close and that there is reason to believe $\sigma_1 = \sigma_2$ (or that the standard deviations are approximately equal). This situation can happen when you make a slight change or alteration to a known process or method of production. The standard deviation may not change much, but the outputs or means could be very different. When there is reason to believe that $\sigma_1 = \sigma_2$, it is best to use a *pooled standard deviation*. The sample test statistic $\bar{x}_1 - \bar{x}_2$ has a corresponding t variable with an *exact* Student's t distribution and degrees of freedom $d.f. = n_1 + n_2 - 2$. Problem 14 at the end of this section provides the details for using the pooled standard deviation in hypothesis tests of and confidence intervals for $\mu_1 - \mu_2$.

Summary

Depending on the information available, there are several methods for testing and constructing confidence intervals for the difference of means $\mu_1 - \mu_2$ from two independent random samples. To use the normal probability distribution, you need to know σ_1 and σ_2. In addition, you need to know that the original population distributions are both normal or that n_1 and n_2 are both of size at least 30. To use a Student's t distribution, you need to know that the original population distributions are both normal or at least mound-shaped, or that n_1 and n_2 are both at least 30. The following decision chart summarizes the appropriate formulas.



TECH NOTE The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all support testing the difference of means for independent samples. The TI-84Plus/TI-83Plus calculators and Minitab also supply confidence intervals for the difference of means.

TI-84Plus/TI-83Plus Enter either summary statistics or raw data. Press **STAT** and select **TESTS.** Options **3:2-SampZTest** and **4:2-SampTTest** test the difference of means using the normal distribution and a Student's t distribution, respectively. Choice **9:2-SampZInt** finds a confidence interval for a difference of means when σ_1 and σ_2 are known. Choice **0:2-SampTInt** finds a confidence interval for a difference of means when σ_1 and σ_2 are unknown. In general, use **No** for Pooled. Then Satterthwaite's approximation for the degrees of freedom is used. However, if $\sigma_1 \approx \sigma_2$, use **Yes** for Pooled.

Excel Enter the data in two columns. Use the menu choices **Tools ➤ Data Analysis.** The choice **z-Test Two Sample Means** tests the difference of means using the normal distribution. The choice **t-Test: Two-Sample Assuming Unequal Variances** tests the difference of means using a Student's t distribution with Satterthwaite's approximation for the degrees of freedom. The choice **t-Test: Two-Sample Assuming Equal Variances** conducts a test using the pooled standard deviation.

Minitab For differences of means, enter the data into two columns. Use the menu choices **STAT ➤ Basic Statistics ➤ 2 sample t.** Minitab always uses a Student's t distribution for $\mu_1 - \mu_2$. In the dialogue box, leave the box **Assume equal variances** unchecked to use Satterthwaite's approximation for the degrees of freedom. Check the box **Assume equal variances** to use the pooled standard deviation.

Testing $\mu_1 - \mu_2$ Using Critical Regions (Optional)

For a set level of significance, the traditional critical region method yields the same results as the P-value method of hypothesis testing. Critical values z_0 for tests using the standard normal distribution can be found in Table 3(c) of the Appendix. Critical values t_0 for tests using a Student's t distribution are found in Table 4 of the Appendix. Use the row headed by the appropriate degrees of freedom and the column that includes the value of α (level of significance) in the *one-tail area row* for one-tailed tests or the *two-tail area row* for two-tailed tests.

EXAMPLE 6

Critical region method

Use the critical region method to solve the application in Example 5 (test $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown).

SOLUTION: Example 5 involves testing the difference in average times for two headache remedies to reach the bloodstream. For brand A, $\bar{x}_1 = 21.8$ min, $s_1 = 8.7$ min, and $n_1 = 12$; for brand B, $\bar{x}_2 = 18.9$ min, $s_2 = 7.5$ min, and $n_2 = 12$. The level of significance α is 0.05. The Student's t distribution is appropriate because both populations are approximately normal.

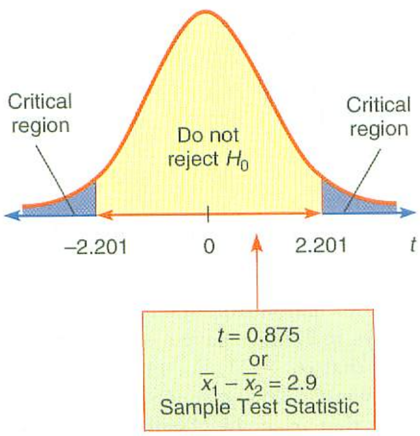
(a) To use the critical region method to test for a difference in average times, we use the same hypotheses and the same sample test statistic as in Example 5.

H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$;

sample test statistic: $\bar{x}_1 - \bar{x}_2 = 2.9$ min; sample $t = 0.875$

FIGURE 10-8

Critical Regions $\alpha = 0.05$; $d.f. = 11$



(b) Instead of finding the P-value of the sample test statistic, we use α and H_1 to find the critical values in Table 4 of the Appendix. We have $d.f. = 11$ (since both samples are of size 12). We find $\alpha = 0.05$ in the *two-tail area* row, since we have a two-tailed test. The critical values are $\pm t_0 = \pm 2.201$.

Next compare the sample test statistic $t = 0.875$ to the critical values. Figure 10-8 shows the critical regions and the sample test statistic. We see that the sample test statistic falls in the "do not reject H_0" region. At the 5% level of significance, the sample evidence does not show a difference in times for the drugs to reach the bloodstream. The result is consistent with the result obtained by the P-value method of Example 5. ◇

VIEWPOINT

Temper! Temper!

In her book *Red Ink Behaviors,* Jean Hollands discusses inappropriate, problem behaviors of professional employees in the corporate business world. Temper tantrums, flaming e-mails, omitting essential information, sabotaging fellow workers, and the arrogant opinion that others are "dumb and dispensable" create personnel problems that cost companies a lot in the form of wasted time, reduced productivity, and lost revenues. A study of major industries in the Silicon Valley area gave Hollands data for estimating just how much time and money are wasted by such "red ink behaviors." For more information, see Problems 11 and 12 in this section.

SECTION 10.2 PROBLEMS

Please provide the following information for Problems 1–12, part (a).

(i) What is the level of significance? State the null and alternate hypotheses.

(ii) What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?

(iii) Find (or estimate) the P-value. Sketch the sampling distribution and show the area corresponding to the P-value.

(iv) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?

(v) State your conclusion in the context of the application.

Note: For degrees of freedom $d.f.$ not in the Student's t table, use the closest $d.f.$ that is *smaller.* In some situations, this choice of $d.f.$ may increase the P-value a small amount, and therefore produce a slightly more "conservative" answer.

Answers may vary due to rounding.

1. *Medical: REM Sleep* REM (rapid eye movement) sleep is sleep during which most dreams occur. Each night a person has both REM and non-REM sleep. However, it is thought that children have more REM sleep than adults (Reference: *Secrets of Sleep* by Dr. A. Borbely). Assume that REM sleep time is normally distributed for both children and adults. A random sample of $n_1 = 10$ children (9 years old) showed that they had an average REM sleep time of $\bar{x}_1 = 2.8$ hours per night. From previous studies it is known that $\sigma_1 = 0.5$ hour. Another random sample of $n_2 = 10$ adults showed that they had an average REM sleep time of $\bar{x}_2 = 2.1$ hours per night. Previous studies show that $\sigma_2 = 0.7$ hour.
   (a) Do these data indicate that on average, children tend to have more REM sleep than adults? Use a 1% level of significance.
   (b) Find a 98% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

2. *Environment: Pollution Index* Based on information from the *Rocky Mountain News*, a random sample of $n_1 = 12$ winter days in Denver gave a sample mean pollution index of $\bar{x}_1 = 43$. Previous studies show that $\sigma_1 = 21$. For Englewood (a suburb of Denver), a random sample of $n_2 = 14$ winter days gave a sample mean pollution index of $\bar{x}_2 = 36$. Previous studies show that $\sigma_2 = 15$. Assume the pollution index is normally distributed in both Englewood and Denver.
   (a) Do these data indicate that the mean population pollution index of Englewood is different (either way) from that of Denver in the winter? Use a 1% level of significance.
   (b) Find a 99% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

3. *Survey: Outdoor Activities* A Michigan study concerning preference for outdoor activities used a questionnaire with a six-point Likert-type response in which 1 designated "not important" and 6 designated "extremely important." A random sample of $n_1 = 46$ adults were asked about fishing as an outdoor activity. The mean response was $\bar{x}_1 = 4.9$. Another random sample of $n_2 = 51$ adults were asked about camping as an outdoor activity. For this group, the mean response was $\bar{x}_2 = 4.3$. From previous studies it is known that $\sigma_1 = 1.5$ and $\sigma_2 = 1.2$. *Note:* A *Likert scale* usually has to do with approval of or agreement with a statement in a questionnaire. For example, respondents are asked to indicate whether they "strongly agree," "agree," "disagree," or "strongly disagree" with the statement.
   (a) Does this indicate a difference (either way) regarding preference for camping versus preference for fishing as an outdoor activity? Use a 5% level of significance.
   (b) Find a 95% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

4. *Generation Gap: Education* Education influences attitude and lifestyle. Differences in education are a big factor in the "generation gap." Is the younger generation really better educated? Large surveys of people age 65 and older were taken in $n_1 = 32$ U.S. cities. The sample mean for these cities showed that $\bar{x}_1 = 15.2\%$ of the older adults had attended college. Large surveys of young adults (age 25–34) were taken in $n_2 = 35$ U.S. cities. The sample mean for these cities showed that $\bar{x}_2 = 19.7\%$ of the young adults had attended college. From previous studies it is known that $\sigma_1 = 7.2\%$ and $\sigma_2 = 5.2\%$ (Reference: *American Generations*, S. Mitchell).
   (a) Does this information indicate that the population mean percentage of young adults who attended college is higher? Use $\alpha = 0.05$.
   (b) Find a 90% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

5. *Crime Rate: FBI* A random sample of $n_1 = 10$ regions in New England gave the following violent crime rates (per million population).

x_1: **New England crime rate**

3.5	3.7	4.0	3.9	3.3	4.1	1.8	4.8	2.9	3.1

Another random sample of $n_2 = 12$ regions in the Rocky Mountain states gave the following violent crime rates (per million population).

x_2: **Rocky Mountain states crime rate**

3.7	4.3	4.5	5.3	3.3	4.8	3.5	2.4	3.1	3.5	5.2	2.8

(Reference: *Crime in the United States,* Federal Bureau of Investigation.) Assume that the crime rate distribution is approximately normal in both regions. Use a calculator to verify that $\bar{x}_1 \approx 3.51$, $s_1 \approx 0.81$, $\bar{x}_2 \approx 3.87$, and $s_2 \approx 0.94$.

(a) Do the data indicate that the violent crime rate in the Rocky Mountain region is higher than in New England? Use $\alpha = 0.01$.

(b) Find a 98% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

6. *Medical: Hay Fever* A random sample of $n_1 = 16$ communities in western Kansas gave the following information for people under 25 years of age.

x_1: **Rate of hay fever per 1000 population for people under 25**

98	90	120	128	92	123	112	93
125	95	125	117	97	122	127	88

A random sample of $n_2 = 14$ regions in western Kansas gave the following information for people over 50 years old.

x_2: **Rate of hay fever per 1000 population for people over 50**

95	110	101	97	112	88	110
79	115	100	89	114	85	96

(Reference: National Center for Health Statistics.)

Use a calculator to verify that $\bar{x}_1 \approx 109.50$, $s_1 \approx 15.41$, $\bar{x}_2 \approx 99.36$, and $s_2 \approx 11.57$.

(a) Assume that the hay fever rate in each age group has an approximately normal distribution. Do the data indicate that the age group over 50 has a lower rate of hay fever? Use $\alpha = 0.05$.

(b) Find a 90% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

7. *Education: Tutoring* In the journal *Mental Retardation,* an article reported the results of a peer tutoring program to help mildly mentally retarded children learn to read. In the experiment, the mildly retarded children were randomly divided into two groups: the experimental group received peer tutoring along with regular instruction, and the control group received regular instruction with no peer tutoring. There were $n_1 = n_2 = 30$ children in each group. The Gates-MacGintie Reading Test was given to both groups before instruction began. For the experimental group, the mean score on the vocabulary portion of the test was $\bar{x}_1 = 344.5$ with sample standard deviation $s_1 = 49.1$. For the control group, the mean score on the same test was $\bar{x}_2 = 354.2$ with sample standard deviation $s_2 = 50.9$.

(a) Use a 5% level of significance to test the hypothesis that there was no difference in the vocabulary scores of the two groups before the instruction began.

(b) Find a 95% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

8. *Education: Tutoring* In the article cited in Problem 7, the results of the following experiment were reported. Form 2 of the Gates-MacGintie Reading Test was

administered to both an experimental group and a control group after 6 weeks of instruction during which the experimental group received peer tutoring and the control group did not. For the experimental group with $n_1 = 30$ children, the mean score on the vocabulary portion of the test was $\bar{x}_1 = 368.4$ with sample standard deviation $s_1 = 39.5$. The average score on the vocabulary portion of the test for the $n_2 = 30$ subjects in the control group was $\bar{x}_2 = 349.2$, with sample standard deviation $s_2 = 56.6$.

(a) Use a 1% level of significance to test the claim that the experimental group performed better than the control group.

(b) Find a 98% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

9. *Wildlife: Fox Rabies* A study of fox rabies in southern Germany gave the following information about different regions and the occurrence of rabies in each region (Reference: B. Sayers, et al., "A Pattern Analysis Study of a Wildlife Rabies Epizootic," *Medical Informatics* 2:11–34). Based on information from this article, a random sample of $n_1 = 16$ locations in region I gave the following information about the number of cases of fox rabies near that location.

x_1: Region I data	1	8	8	8	7	8	8	1
	3	3	3	2	5	1	4	6

A second random sample of $n_2 = 15$ locations in region II gave the following information about the number of cases of fox rabies near that location.

x_2: Region II data	1	1	3	1	4	8	5	4
	4	4	2	2	5	6	9	

Use a calculator with sample mean and sample standard deviation keys to verify that $\bar{x}_1 = 4.75$ with $s_1 \approx 2.82$ in region I and $\bar{x}_2 \approx 3.93$ with $s_2 \approx 2.43$ in region II.

(a) Does this information indicate that there is a difference (either way) in the mean number of cases of fox rabies between the two regions? Use a 5% level of significance. (Assume the distribution of rabies cases in both regions is mound-shaped and approximately normal.)

(b) Find a 95% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

10. *Agriculture: Bell Peppers* The pathogen *Phytophthora capsici* causes bell peppers to wilt and die. Because bell peppers are an important commercial crop, this disease has undergone a great deal of agricultural research. It is thought that too much water aids the spread of the pathogen. Two fields are under study. The first step in the research project is to compare the mean soil water content for the two fields (Source: *Journal of Agricultural, Biological, and Environmental Statistics,* Vol. 2, No. 2). Units are percent water by volume of soil.

Field A samples, x_1:

10.2	10.7	15.5	10.4	9.9	10.0	16.6
15.1	15.2	13.8	14.1	11.4	11.5	11.0

Field B samples, x_2:

8.1	8.5	8.4	7.3	8.0	7.1	13.9	12.2
13.4	11.3	12.6	12.6	12.7	12.4	11.3	12.5

Use a calculator with mean and standard deviation keys to verify that $\bar{x}_1 \approx 12.53$, $s_1 \approx 2.39$, $\bar{x}_2 \approx 10.77$, and $s_2 \approx 2.40$.

(a) Assuming the distribution of soil water content in each field is mound-shaped and symmetric, use a 5% level of significance to test the claim that field A has, on average, a higher soil water content than field B.

(b) Find a 90% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

11. *Management: Lost Time* In her book *Red Ink Behaviors*, Jean Hollands reports on the assessment of leading Silicon Valley companies regarding a manager's lost time due to inappropriate behavior of employees. Consider the following independent random variables. The first variable x_1 measures manager's hours per week lost due to hot tempers, flaming e-mails, and general unproductive tensions:

x_1:	1	5	8	4	2	4	10

The variable x_2 measures manager's hours per week lost due to disputes regarding technical workers' superior attitudes that their colleagues are "dumb and dispensable":

x_2:	10	5	4	7	9	4	10	3

Use a calculator with sample mean and standard deviation keys to verify that $\bar{x}_1 \approx 4.86$, $s_1 \approx 3.18$, $\bar{x}_2 = 6.5$, and $s_2 \approx 2.88$.

(a) Does the information indicate that the population mean time lost due to hot tempers is different (either way) from the population mean time lost due to disputes arising from technical workers' superior attitudes? Use $\alpha = 0.05$. Assume that the two lost-time population distributions are mound-shaped and symmetric.

(b) Find a 95% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

12. *Management: Intimidators and Stressors* This problem is based on information regarding productivity in leading Silicon Valley companies (see reference in Problem 11). In large corporations, an "intimidator" is an employee who tries to stop communication, sometimes sabotages others, and, above all, likes to listen to him- or herself talk. Let x_1 be a random variable representing productive hours per week lost by peer employees of an intimidator.

x_1:	8	3	6	2	2	5	2

A "stressor" is an employee with a hot temper that leads to unproductive tantrums in corporate society. Let x_2 be a random variable representing productive hours per week lost by peer employees of a stressor.

x_2:	3	3	10	7	6	2	5	8

Use a calculator with mean and standard deviation keys to verify that $\bar{x}_1 = 4.00$, $s_1 \approx 2.38$, $\bar{x}_2 = 5.5$, and $s_2 \approx 2.78$.

(a) Assuming that the variables x_1 and x_2 are independent, do the data indicate that the population mean time lost due to stressors is greater than the population mean time lost due to intimidators? Use a 5% level of significance. (Assume that the population distributions of time lost due to intimidators and time lost due to stressors are each mound-shaped and symmetric.)

(b) Find a 90% confidence interval for $\mu_1 - \mu_2$. Explain the meaning of the confidence interval in the context of the problem.

13. *Expand Your Knowledge: Software Approximation for Degrees of Freedom* Given x_1 and x_2 distributions that are normal or approximately normal with unknown σ_1 and σ_2, the value of t corresponding to $\bar{x}_1 - \bar{x}_2$ has a distribution that is approximated by a Student's t distribution. We use the convention that the degrees of freedom

is approximately the smaller of $n_1 - 1$ and $n_2 - 1$. However, a more accurate estimate for the appropriate degrees of freedom is given by Satterthwaite's formula:

$$d.f. \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

where s_1, s_2, n_1, and n_2 are the respective sample standard deviations and sample sizes of independent random samples from the x_1 and x_2 distributions. This is the approximation used by most statistical software. When both n_1 and n_2 are 5 or larger, it is quite accurate. The degrees of freedom computed from this formula are either truncated or not rounded.

(a) In Problem 5 we tested whether the population average crime rate μ_2 in the Rocky Mountain region is higher than that in New England, μ_1. The data were $n_1 = 10$, $\bar{x}_1 \approx 3.51$, $s_1 \approx 0.81$, $n_2 = 12$, $\bar{x}_2 \approx 3.87$, and $s_2 \approx 0.94$. Use Satterthwaite's formula to compute the degrees of freedom for the Student's t distribution.

(b) When you did Problem 5, you followed the convention that degrees of freedom $d.f.$ = *smaller* of $n_1 - 1$ and $n_2 - 1$. Compare this $d.f.$ with that found by Satterthwaite's formula.

14. *Expand Your Knowledge: Pooled Two-Sample Procedure* Consider independent random samples from two populations that are normal or approximately normal, or the case in which both sample sizes are at least 30. Then, if σ_1 and σ_2 are unknown but we have reason to believe that $\sigma_1 = \sigma_2$, we can pool the standard deviations. Using sample sizes n_1 and n_2, the sample test statistic $\bar{x}_1 - \bar{x}_2$ has a Student's t distribution, where

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with degrees of freedom } d.f. = n_1 + n_2 - 2$$

where the **pooled standard deviation** s is

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Hypothesis tests: Use H_0: $\mu_1 = \mu_2$ and an appropriate alternate hypothesis. The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } d.f. = n_1 + n_2 - 2$$

A c confidence interval: $(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$

where $E = t_c\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

t_c = critical value for confidence level c and degrees of freedom $d.f. = n_1 + n_2 - 2$

Note: With statistical software, select the pooled variance or equal variance options.

There are many situations in which we want to compare means from populations having standard deviations that are equal. This method applies even if the standard deviations are known to be only approximately equal. Consider Problem 9 regarding average incidence of fox rabies in two regions. For region I, $n_1 = 16$, $\bar{x}_1 = 4.75$, and $s_1 \approx 2.82$ and for region II, $n_2 = 15$, $\bar{x}_2 \approx 3.93$, and $s_2 \approx 2.43$. The two sample standard deviations are sufficiently close that we can assume $\sigma_1 = \sigma_2$.
(a) Use the method of pooled standard deviation to redo Problem 9(a).
(b) Use the method of pooled standard deviation to redo Problem 9(b).

15. *Critical Region Method: Testing $\mu_1 - \mu_2$; σ_1, σ_2 Unknown* Redo Problem 5(a) using the critical region method and compare your results to those obtained using the P-value method.

16. *Critical Region Method: Testing $\mu_1 - \mu_2$; σ_1, σ_2 Known* Redo Problem 1(a) using the critical region method and compare your results to those obtained using the P-value method.

10.3 Inferences About the Difference of Two Proportions $p_1 - p_2$

FOCUS POINTS

- Compute the sample test statistic and P-value for testing $p_1 - p_2$.
- Find confidence intervals for $p_1 - p_2$.

Testing a Difference of Proportions $p_1 - p_2$

Is the population proportion of people favoring more wilderness areas different between men and women, between younger people and older people, between ranchers and energy employees, etc.? To answer questions of this type, we explore the sampling distribution $\hat{p}_1 - \hat{p}_2$ from two independent binomial experiments.

THEOREM 10.3 Suppose we have two independent binomial experiments. That is, outcomes from one binomial experiment are in no way paired with outcomes from the other. We use the notation

Binomial Experiment 1	*Binomial Experiment 2*
n_1 = number of trials	n_2 = number of trials
r_1 = number of successes	r_2 = number of successes
p_1 = population probability of success on a single trial	p_2 = population probability of success on a single trial

For *large* values of n_1 and n_2, the distribution of sample differences

$$\hat{p}_1 - \hat{p}_2 = \frac{r_1}{n_1} - \frac{r_2}{n_2}$$

is closely approximated by a *normal distribution* with mean μ and standard deviation σ as shown:

$$\mu = p_1 - p_2 \qquad \sigma = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

Pooled estimate $\bar{p}$

For most practical problems involving a comparison of two binomial populations, the experimenters will want to test the null hypothesis $p_1 = p_2$. Consequently, this is the type of test we shall consider. Since the values of p_1 and p_2 are unknown, and since specific values are not assumed under the null hypothesis $p_1 = p_2$, the best estimate for the common value is the total number of successes $(r_1 + r_2)$ divided by the total number of trials $(n_1 + n_2)$. If we denote this *pooled estimate of proportion* by $\bar{p}$ (read "p bar"), then

$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2}$$

This formula gives the best sample estimate $\bar{p}$ for p_1 and p_2 *under the assumption that* $p_1 = p_2$. Also, $\bar{q} = 1 - \bar{p}$.

Criteria for using the normal approximation to the binomial

◇ **COMMENT** For most practical applications, the sample sizes n_1 and n_2 are considered large samples if each of the four quantities

$$n_1\bar{p} \quad n_1\bar{q} \quad n_2\bar{p} \quad n_2\bar{q}$$

is larger than 5 (see Section 7.6). ◇

Theorem 10.3 leads to the following procedure for testing $p_1 - p_2$.

PROCEDURE

How to test a difference of proportions $p_1 - p_2$

Consider two independent binomial experiments.

Binomial Experiment 1	*Binomial Experiment 2*
n_1 = number of trials	n_2 = number of trials
r_1 = number of successes out of n_1 trials	r_2 = number of successes out of n_2 trials
$\hat{p}_1 = \frac{r_1}{n_1}$	$\hat{p}_2 = \frac{r_2}{n_2}$
p_1 = population probability of success on a single trial	p_2 = population probability of success on a single trial

1. Use the *null hypothesis* of no difference, H_0: $p_1 - p_2 = 0$. In the context of the application, choose the *alternate hypothesis*. Set the *level of significance* α.
2. The null hypothesis claims that $p_1 = p_2$; therefore, *pooled best estimates* for the population probabilities of success and failure are

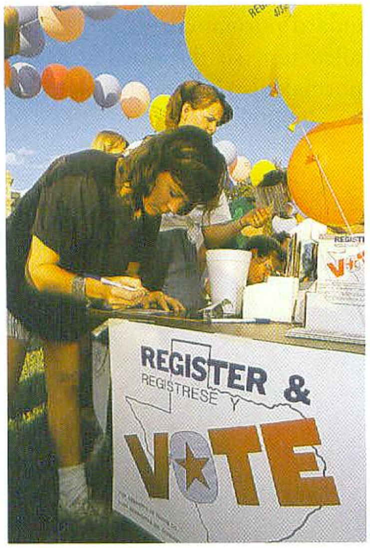
$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2} \quad \text{and} \quad \bar{q} = 1 - \bar{p}$$

The number of trials should be sufficiently large so that all four quantities $n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, and $n_2\bar{q}$ are each larger than 5. In this case you compute the sample *test statistic*

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

3. Use the standard normal distribution and a type of test, one-tailed or two-tailed, to find the *P-value* corresponding to the sample 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. *State your conclusion* in the context of the application.

EXAMPLE 7
Testing the difference of proportions



The Macek County Clerk wishes to improve voter registration. One method under consideration is to send reminders in the mail to all citizens in the county who are eligible to register. As part of a pilot study to determine if this method will actually improve voter registration, a random sample of 1250 potential voters was taken. This sample was then randomly divided into two groups.

Group 1: There were 625 people in this group. No reminders to register were sent to them. The number of potential voters from this group who registered was 295.

Group 2: This group also contained 625 people. Reminders were sent in the mail to each member in the group, and the number who registered to vote was 350.

The county clerk claims that the proportion of people who registered was significantly greater in group 2. On the basis of this claim, the clerk recommends that the project be funded for the entire population of Macek County. Use a 5% level of significance to test the claim that the proportion of potential voters who registered was greater in group 2, the group that received reminders.

SOLUTION:

(a) Note that $\alpha = 0.05$. Let p_1 be the proportion of voters who registered from group 1, and let p_2 be the proportion who registered from group 2. The null hypothesis is that there is no difference in proportions, so

$$H_0\text{: } p_1 = p_2 \qquad \text{or} \qquad H_0\text{: } p_1 - p_2 = 0$$

The alternate hypothesis is that the proportion of voters who registered is greater for the group that received reminders.

$$H_1\text{: } p_1 < p_2 \qquad \text{or} \qquad H_1\text{: } p_1 - p_2 < 0$$

(b) Compute the sample statistic $\hat{p}_1 - \hat{p}_2$ and convert it to a z value.

CALCULATOR NOTE Carry the values for $\hat{p}_1$, $\hat{p}_2$, and the pooled estimates $\bar{p}$ and $\bar{q}$ to at least three places after the decimal. Then round the z value of the corresponding test statistic to two places after the decimal.

For the first group, the number of successes is $r_1 = 295$ out of $n_1 = 625$ trials. For the second group, there are $r_2 = 350$ successes out of $n_2 = 625$ trials. Since

$$\hat{p}_1 = \frac{r_1}{n_1} = \frac{295}{625} = 0.472 \qquad \text{and} \qquad \hat{p}_2 = \frac{r_2}{n_2} = \frac{350}{625} = 0.560$$

then

$$\hat{p}_1 - \hat{p}_2 = 0.472 - 0.560 = -0.088$$

To convert this $\hat{p}_1 - \hat{p}_2$ value to a z value, we need to find the *pooled estimate* $\bar{p}$ for the common values of p_1 and p_2 and the corresponding value for $\bar{q}$.

$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2} = \frac{295 + 350}{625 + 625} = 0.516 \quad \text{and} \quad \bar{q} = 1 - \bar{p} = 0.484$$

Using these values, we find that

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{-0.088}{\sqrt{\frac{(0.516)(0.484)}{625} + \frac{(0.516)(0.484)}{625}}} \approx -3.11$$

(c) Find the *P*-value and sketch the area on the standard normal curve.
Figure 10-9 shows the *P*-value. This is a left-tailed test, so the *P*-value is the area to the left of -3.11. Using the standard normal distribution (Table 3 of the Appendix), we find *P*-value $= P(z < -3.11) \approx 0.0009$.

FIGURE 10-9

P-value

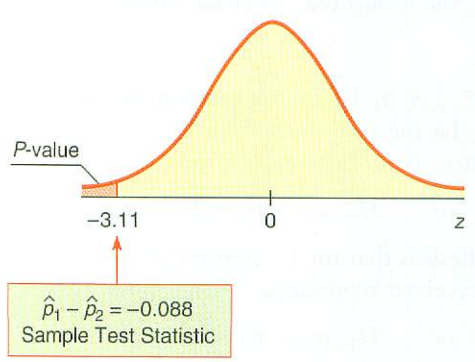
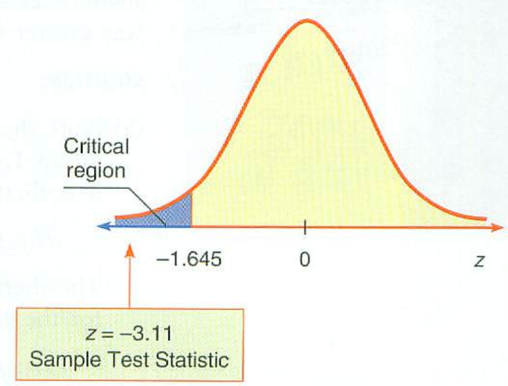


FIGURE 10-10

Critical Region, $\alpha = 0.05$



(d) Conclude the test.
Since *P*-value of $0.0009 \leq 0.05$ for α, we reject H_0.

(e) Interpret the results.
At the 5% level of significance, the data indicate that the population proportion of potential voters who registered was greater in group 2, the group that received reminders.

(f) **Critical region method (optional):** Use the critical region method to conclude the test at the 5% level of significance. Compare your results with the *P*-value method.

In part (b), we found that the sample test statistic is $z = -3.11$. In Table 3(c) of the Appendix, we see that the critical value $z_0 = -1.645$ for a left-tailed test with $\alpha = 0.05$. In Figure 10-10, we see that the sample test statistic falls in the critical region, so we reject H_0. This conclusion is consistent with the conclusion obtained using the *P*-value method. There is sufficient evidence to conclude that at the 5% level of significance, the population of potential voters who registered was greater in the group receiving reminders. ◆

GUIDED EXERCISE 6

Testing the difference of proportions

In Example 7 about voter registration, suppose that a random sample of 1100 potential voters was randomly divided into two groups.

Group 1: 500 potential voters; no registration reminders sent; 248 registered to vote

Group 2: 600 potential voters; registration reminders sent; 332 registered to vote

Do these data support the claim that the proportion of voters who registered was greater in the group that received reminders than in the group that did not? Use a 1% level of significance.

(a) What is α? State H_0 and H_1.

→ $\alpha = 0.01$. As before, H_0: $p_1 = p_2$ and H_1: $p_1 < p_2$.

(b) Under the null hypothesis $p_1 = p_2$, calculate the *pooled estimates* $\bar{p}$ and $\bar{q}$.

→ $n_1 = 500,\ r_1 = 248;\ n_2 = 600,\ r_2 = 332$

$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2} = \frac{248 + 332}{500 + 600} \approx 0.527$$

$$\bar{q} = 1 - \bar{p} \approx 1 - 0.527 \approx 0.473$$

(c) What is the value of the sample test statistic $\hat{p}_1 - \hat{p}_2$?

→ $$\hat{p}_1 = \frac{r_1}{n_1} = \frac{248}{500} = 0.496 \qquad \hat{p}_2 = \frac{r_2}{n_2} = \frac{332}{600} \approx 0.553$$

$$\hat{p}_1 - \hat{p}_2 = -0.057$$

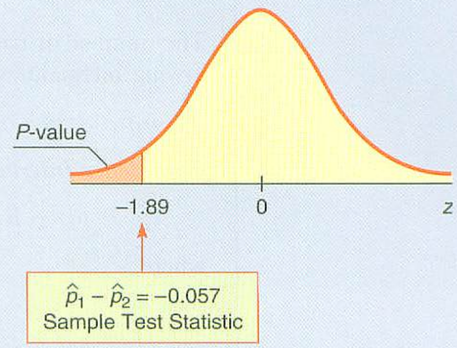
(d) Convert the sample test statistic $\hat{p}_1 - \hat{p}_2 = -0.057$ to a z value.

→ $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}} = \frac{-0.057}{\sqrt{\dfrac{(0.527)(0.473)}{500} + \dfrac{(0.527)(0.473)}{600}}} \approx -1.89$$

(e) Find the P-value and sketch the area on the standard normal curve.

→ Figure 10-11 shows the P-value. It is the area to the left of $z = -1.89$. Using Table 3 of the Appendix, we find P-value $= P(z < -1.89) = 0.0294$.

FIGURE 10-11 P-value



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

→ Since P-value of $0.0294 > 0.01$ for α, we cannot reject H_0. At the 1% level of significance, the data do not support the claim that the reminders increase the proportion of registered voters.

Estimating a Difference of Proportions $p_1 - p_2$

We conclude this section with a discussion of confidence intervals for $p_1 - p_2$, the difference of two proportions from two independent binomial probability distributions. Theorem 10.3 gives us the basis for constructing such confidence intervals. Using the notation of Theorem 10.3, recall that for sufficiently large sample sizes n_1 and n_2, the sampling distribution $\hat{p}_1 - \hat{p}_2$ of differences of proportions from two independent binomial probability distributions is approximately normal, with mean $\mu = p_1 - p_2$ and standard deviation $\sigma = \sqrt{p_1q_1/n_1 - p_2q_2/n_2}$.

There is one technical difficulty in computing the standard deviation σ of the $\hat{p}_1 - \hat{p}_2$ distribution. We don't know the values of p_1 or p_2. However, if all four quantities $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$, and $n_2\hat{q}_2$ are greater than 5, then a good approximation for the standard deviation σ of the $\hat{p}_1 - \hat{p}_2$ distribution is given by

$$\sigma \approx \hat{\sigma} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

This leads to the following procedure for confidence intervals.

PROCEDURE

How to find a confidence interval for $p_1 - p_2$

Consider two independent binomial experiments.

Binomial Experiment 1	*Binomial Experiment 2*
n_1 = number of trials	n_2 = number of trials
r_1 = number of successes out of n_1 trials	r_2 = number of successes out of n_2 trials
$\hat{p}_1 = \frac{r_1}{n_1}$; $\hat{q}_1 = 1 - \hat{p}_1$	$\hat{p}_2 = \frac{r_2}{n_2}$; $\hat{q}_2 = 1 - \hat{p}_2$
p_1 = population probability of success	p_2 = population probability of success

The number of trials should be sufficiently large so that all four of the following inequalities are true:

$$n_1\hat{p}_1 > 5; \quad n_1\hat{q}_1 > 5; \quad n_2\hat{p}_2 > 5; \quad n_2\hat{q}_2 > 5$$

Confidence interval for $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) - E \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + E$$

where

$$E \approx z_c\hat{\sigma} = z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

c = confidence level, $0 < c < 1$

z_c = critical value for confidence level c based on the standard normal distribution (See Table 3(b) of the Appendix for commonly used values.)

EXAMPLE 8

Confidence interval for $p_1 - p_2$

In his book *Secrets of Sleep*, Professor Borbely describes research on dreams in the sleep laboratory at the University of Zurich Medical School. During normal sleep, there is a phase known as *REM* (rapid eye movement). For most people, REM sleep occurs about every 90 minutes or so, and it is thought that dreams occur just before or during the REM phase. Using electronic equipment in the sleep laboratory, it is possible to detect the REM phase in a sleeping person. If a person is wakened immediately after the REM phase, he or she usually can describe a dream that has just taken place. Based on a study of over 650 people in the Zurich sleep laboratory, it was found that about one-third of all dream reports contain feelings of fear, anxiety, or aggression. There is a conjecture that if a person is in a good mood when going to sleep, the proportion of "bad" dreams (fear, anxiety, aggression) might be reduced.

Suppose that two groups of subjects were randomly chosen for a sleep study. In group I, before going to sleep, the subjects spent 1 hour watching a comedy movie. In this group, there were a total of $n_1 = 175$ dreams recorded, of which $r_1 = 49$ were dreams with feelings of anxiety, fear, or aggression. In group II, the subjects did not watch a movie but simply went to sleep. In this group, there were a total of $n_2 = 180$ dreams recorded, of which $r_2 = 63$ were dreams with feelings of anxiety, fear, or aggression.

(a) Why could groups I and II be considered independent binomial distributions? Why do we have a "large-sample" situation?

SOLUTION: Since the two groups were chosen randomly, it is reasonable to assume that neither group's response would be related to the other's. In both groups, each recorded dream could be thought of as a trial, with success being a dream with feelings of fear, anxiety, or aggression.

$$\hat{p}_1 = \frac{r_1}{n_1} = \frac{49}{175} = 0.28 \quad \text{and} \quad \hat{q}_1 = 1 - \hat{p}_1 = 0.72$$

$$\hat{p}_2 = \frac{r_2}{n_2} = \frac{63}{180} = 0.35 \quad \text{and} \quad \hat{q}_2 = 1 - \hat{p}_2 = 0.65$$

Since

$$n_1\hat{p}_1 = 49 > 5 \qquad n_1\hat{q}_1 = 126 > 5$$
$$n_2\hat{p}_2 = 63 > 5 \qquad n_2\hat{q}_2 = 117 > 5$$

then large-sample theory is appropriate.

(b) What is $p_1 - p_2$? Compute a 95% confidence interval for $p_1 - p_2$.

SOLUTION: p_1 is the population proportion of successes (bad dreams) for all people who watch comedy movies before bed. Thus, p_1 can be thought of as the percentage of bad dreams for all people who are in a "good mood" when they go to bed. Likewise, p_2 is the percentage of bad dreams for the population of all people who just go to bed (no movie). The difference $p_1 - p_2$ is the population difference.

To find a confidence interval for $p_1 - p_2$, we need the values of z_c, $\hat{\sigma}$, and then E. From Table 3(b) of the Appendix, we see that $z_{0.95} = 1.96$, so

$$\hat{\sigma} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = \sqrt{\frac{(0.28)(0.72)}{175} + \frac{(0.35)(0.65)}{180}}$$

$$\approx \sqrt{0.0024} \approx 0.0492$$

$$E = z_c\hat{\sigma} = 1.96(0.0492) \approx 0.096$$

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.28 - 0.35) - 0.096 < p_1 - p_2 < (0.28 - 0.35) + 0.096$$

$$-0.166 < p_1 - p_2 < 0.026$$

(c) Explain the meaning of the confidence interval that you constructed in part (b).

SOLUTION: We are 95% sure that the interval between −16.6% and 2.6% is one that contains the percentage difference of "bad" dreams for group I and group II. Since the interval −0.166 to 0.026 is not all negative (or all positive), we cannot say that $p_1 - p_2 < 0$ (or $p_1 - p_2 > 0$). Thus, at the 95% confidence level, we *cannot* conclude that $p_1 < p_2$ or $p_1 > p_2$. The comedy movies before bed help some people reduce the percentage of "bad" dreams, but at the 95% confidence level, we cannot say that the *population difference* is reduced. ◆

TECH NOTE The TI-84Plus and TI-83Plus calculators and Minitab support both testing and confidence intervals for the difference of proportions.

TI-84Plus/TI-83Plus Use the **STAT** key and highlight **TESTS.** The menu choices **B:2-PropZInt** and **6:2-PropZTest** provide confidence intervals and testing, respectively. Note that the symbol x is used to designate the number of successes r.

Minitab Use the menu choices **STAT ➤ Basic Statistics ➤ 2 proportions.** In the dialogue box, under Options, select the null and alternate hypotheses and set the confidence level.

VIEWPOINT

What's the Difference?

Will two 15-minute piano lessons a week significantly improve a child's analytical reasoning skills? Why piano? Why not computer keyboard instruction or maybe voice lessons? Professor Frances Rauscher, University of Wisconsin, and Professor Gordon Shaw, University of California at Irvine, claim that there is a difference! How could this be measured? A large number of piano students were given complicated tests of mental ability. Independent control groups of other students were given the same tests. Techniques involving the study of differences of means and proportions were used to draw the conclusion that students taking piano lessons did better on tests measuring analytical reasoning skills (Reported in *The Denver Post*).

SECTION 10.3 PROBLEMS

Please provide the following information for Problems 1–8.

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

(b) What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?

(c) Find the *P*-value. Sketch the sampling distribution and show the area corresponding to the *P*-value.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
(e) State your conclusion in the context of the application.

Note: Answers may vary due to rounding.

1. *Federal Tax Money: Art Funding* Would you favor spending more federal tax money on the arts? This question was asked by a research group on behalf of *The National Institute* (Reference: *Painting by Numbers,* J. Wypijewski, University of California Press). Of a random sample of $n_1 = 220$ women, $r_1 = 59$ responded yes. Another random sample of $n_2 = 175$ men showed that $r_2 = 56$ responded yes. Does this information indicate a difference (either way) between the population proportion of women and the population proportion of men who favor spending more federal tax dollars on the arts? Use $\alpha = 0.05$.

2. *Art Funding: Politics* Would you favor spending more federal tax money on the arts? This question was asked by a research group on behalf of *The National Institute* (Reference: *Painting by Numbers,* J. Wypijewski, University of California Press). Of a random sample of $n_1 = 93$ politically conservative voters, $r_1 = 21$ responded yes. Another random sample of $n_2 = 83$ politically moderate voters showed that $r_2 = 22$ responded yes. Does this information indicate that the population proportion of conservative voters inclined to spend more federal tax money on funding the arts is less than the proportion of moderate voters so inclined? Use $\alpha = 0.05$.

3. *Sociology: Trusting People* Generally speaking, would you say that most people can be trusted? A random sample of $n_1 = 250$ people in Chicago ages 18–25 showed that $r_1 = 45$ said yes. Another random sample of $n_2 = 280$ people in Chicago ages 35–45 showed that $r_2 = 71$ said yes (based on information from the *National Opinion Research Center,* University of Chicago). Does this indicate that the population proportion of trusting people in Chicago is higher for the older group? Use $\alpha = 0.05$.

4. *Political Science: Voters* This problem is based on information taken from *Life in America's Fifty States,* by G. S. Thomas. A random sample of $n_1 = 288$ voters registered in the state of California showed that 141 voted in the last general election. A random sample of $n_2 = 216$ registered voters in the state of Colorado showed that 125 voted in the most recent general election. Do these data indicate that the population proportion of voter turnout in Colorado is higher than that in California? Use a 5% level of significance.

5. *Extraterrestrials: Believe It?* Based on information from *Harper's Index,* $r_1 = 37$ out of a random sample of $n_1 = 100$ adult Americans who did not attend college believe in extraterrestrials. However, out of a random sample of $n_2 = 100$ adult Americans who did attend college, $r_2 = 47$ claim that they believe in extraterrestrials. Does this indicate that the proportion of people who attended college and who believe in extraterrestrials is higher than the proportion who did not attend college? Use $\alpha = 0.01$.

6. *Art: Politics* Do you prefer paintings in which the people are fully clothed? This question was asked by a professional survey group on behalf of the National Arts Society (see reference in Problem 2). A random sample of $n_1 = 59$ people who are conservative voters showed that $r_1 = 45$ said yes. Another random sample of $n_2 = 62$ people who are liberal voters showed that $r_2 = 36$ said yes. Does this indicate that the population proportion of conservative voters who prefer art with fully clothed people is higher? Use $\alpha = 0.05$.

7. *Hotels: Nonsmoking* A random sample of $n_1 = 378$ hotel guests was taken 1 year ago, and it was found that $r_1 = 194$ requested nonsmoking rooms. Recently, a random sample of $n_2 = 516$ hotel guests showed that $r_2 = 320$ requested nonsmoking rooms. Do these data indicate that the proportion of hotel guests requesting nonsmoking rooms has increased? Use a 1% level of significance.

8. *Sociology: College Degrees* A random sample of $n_1 = 78$ women ages 21–29 in Denver showed that $r_1 = 23$ have a college degree. Another random sample of $n_2 = 73$ men in Denver in the same age group showed that $r_2 = 20$ have a college degree (based on information from *Educational Attainment in the United States,* Bureau of the Census). Does this indicate that the population proportion of Denver women ages 21–29 with college degrees is different (either way) from that of men in this age group? Use $\alpha = 0.05$.

9. *Critical Region Method: Testing $p_1 - p_2$* Redo Problem 1 using the critical region method and compare your results to those obtained using the P-value method.

10. *Critical Region Method: Testing $p_1 - p_2$* Redo Problem 2 using the critical region method and compare your results to those obtained using the P-value method.

11. *Myers-Briggs: Marriage Counseling* Isabel Myers was a pioneer in the study of personality types. She identified four basic personality preferences that are described at length in the book *A Guide to the Development and Use of the Myers-Briggs Type Indicator,* by Myers and McCaulley (Consulting Psychologists Press). Marriage counselors know that couples who have none of the four preferences in common may have a stormy marriage. Myers took a random sample of 375 married couples and found that 289 had two or more personality preferences in common. In another random sample of 571 married couples, it was found that only 23 had no preferences in common. Let p_1 be the population proportion of all married couples who have two or more personality preferences in common. Let p_2 be the population proportion of all married couples who have no personality preferences in common.
    (a) Find a 99% confidence interval for $p_1 - p_2$.
    (b) Explain the meaning of the confidence interval in part (a) in the context of this problem. Does the confidence interval contain all positive, all negative, or both positive and negative numbers? What does this tell you (at the 99% confidence level) about the proportion of married couples with two or more personality preferences in common compared with the proportion of married couples sharing no personality preferences in common?

12. *Myers-Briggs: Marriage Counseling* Most married couples have two or three personality preferences in common (see reference in Problem 11). Myers used a random sample of 375 married couples and found that 132 had three preferences in common. Another random sample of 571 couples showed that 217 had two personality preferences in common. Let p_1 be the population proportion of all married couples who have three personality preferences in common. Let p_2 be the population proportion of all married couples who have two personality preferences in common.
    (a) Find a 90% confidence interval for $p_1 - p_2$.
    (b) Examine the confidence interval in part (a) and explain what it means in the context of this problem. Does the confidence interval contain all positive, all negative, or both positive and negative numbers? What does this tell you about the proportion of married couples with three personality preferences in common compared with the proportion of couples with two preferences in common (at the 90% confidence level)?

13. *Navajo Culture: Traditional Hogans* S. C. Jett is a professor of geography at the University of California, Davis. He and a colleague, V. E. Spencer, are experts on modern Navajo culture and geography. The following information is taken from

their book *Navajo Architecture: Forms, History, Distributions* (University of Arizona Press). On the Navajo Reservation, a random sample of 210 permanent dwellings in the Fort Defiance region showed that 65 were traditional Navajo hogans. In the Indian Wells region, a random sample of 152 permanent dwellings showed that 18 were traditional hogans. Let p_1 be the population proportion of all traditional hogans in the Fort Defiance region, and let p_2 be the population proportion of all traditional hogans in the Indian Wells region.
(a) Find a 99% confidence interval for $p_1 - p_2$.
(b) Examine the confidence interval and comment on its meaning. Does it include numbers that are all positive? all negative? mixed? What if it is hypothesized that Navajo who follow the traditional culture of their people tend to occupy hogans? Comment on the confidence interval for $p_1 - p_2$ in this context.

14. *Archaeology: Cultural Affiliation* "Unknown cultural affiliations and loss of identity at high elevations." These are words used to propose the hypothesis that archaeological sites tend to lose their identity as altitude extremes are reached. This idea is based on the notion that prehistoric people tended *not* to take trade wares to temporary settings and/or isolated areas (Source: *Prehistoric New Mexico: Background for Survey*, by D. E. Stuart and R. P. Gauthier, University of New Mexico Press). As elevation zones of prehistoric people (in what is now the state of New Mexico) increased, there seemed to be a loss of artifact identification. Consider the following information.

Elevation Zone	Number of Artifacts	Number Unidentified
7000–7500 ft	112	69
5000–5500 ft	140	26

Let p_1 be the population proportion of unidentified archaeological artifacts at the elevation zone 7000–7500 ft in the given archaeological area. Let p_2 be the population proportion of unidentified archaeological artifacts at the elevation zone 5000–5500 ft in the given archaeological area.
(a) Find a 99% confidence interval for $p_1 - p_2$.
(b) Explain the meaning of the confidence interval in part (a) in the context of this problem. Does the confidence interval contain all positive numbers? all negative numbers? both positive and negative numbers? What does this tell you (at the 99% confidence level) about the population proportion of unidentified artifacts at high elevations (7000–7500 ft) compared with the population proportion of unidentified artifacts at lower elevations (5000–5500 ft)? How does this relate to the stated hypothesis?

15. *General: Different Confidence Levels*
(a) Suppose that a 95% confidence interval for a difference of proportions contains both positive and negative numbers. Will a 99% confidence interval based on the same data necessarily contain both positive and negative numbers? Explain. What about a 90% confidence interval? Explain.
(b) Suppose that a 95% confidence interval for a difference of proportions contains all positive numbers. Will a 99% confidence interval based on the same data necessarily contain all positive numbers as well? Explain. What about a 90% confidence interval? Explain.

16. *Focus Problem: Wood Duck Nests* In the Focus Problem at the beginning of this chapter, a study was described comparing the hatch ratios of wood duck nesting boxes. Group I nesting boxes were well separated from each other and well hidden by available brush. There were a total of 474 eggs in group I boxes, of which a field count

showed about 270 hatched. Group II nesting boxes were placed in highly visible locations and grouped closely together. There were a total of 805 eggs in group II boxes, of which a field count showed about 270 hatched.

(a) Find a point estimate $\hat{p}_1$ for p_1, the proportion of eggs that hatch in group I nest box placements. Find a 95% confidence interval for p_1.

(b) Find a point estimate $\hat{p}_2$ for p_2, the proportion of eggs that hatch in group II nest box placements. Find a 95% confidence interval for p_2.

(c) Find a 95% confidence interval for $p_1 - p_2$. Does the interval indicate that the proportion of eggs hatched from group I nest boxes is higher than, lower than, or equal to the proportion of eggs hatched from group II nest boxes?

(d) What conclusions about placement of nest boxes can be drawn? In the article discussed in the Focus Problem, additional concerns are raised about the higher cost of placing and maintaining group I nest boxes. Also at issue is the cost efficiency per successful wood duck hatch.

SUMMARY

In this chapter, we continued our discussion of hypothesis testing and confidence intervals. We considered paired data from dependent samples and tested the population mean μ_d of differences between the members of the data pairs. For two independent samples, we considered the population difference of means $\mu_1 - \mu_2$, where μ_1 is the population mean of the first sample and μ_2 is that of the second sample. Hypothesis tests and confidence intervals were developed for the difference of means. The sample statistic $\bar{x}_1 - \bar{x}_2$ follows a normal or Student's t distribution, depending on the given conditions and whether the population standard deviations σ_1 and σ_2 are known. Appropriate methods of testing and confidence intervals were developed for each case.

Finally, we looked at the population difference of proportions $p_1 - p_2$ from two large, independent samples. We studied both hypothesis testing and confidence intervals for the difference of proportions.

IMPORTANT WORDS & SYMBOLS

Section 10.1

Dependent samples

Data pairs

μ_d, difference of means from data pairs

Section 10.2

Independent samples

d.f. for testing $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown

Pooled standard deviation

Section 10.3

Pooled estimate of proportion $\bar{p}$

VIEWPOINT

Who Watches Cable TV?

Consider the following claim: Average cable TV viewers are, generally speaking, as affluent as newspaper readers and better off than radio or magazine audiences. How do we know that this claim is true? One way to answer such a question is to construct several individual tests of difference of means. One test would compare cable TV viewers with newspaper readers, a second test would compare cable TV viewers with radio audiences, and a third test would compare cable TV viewers with magazine audiences. Another way to handle all three tests at once would be to use techniques of ANOVA (analysis of variance)—a subject often introduced in a second course in statistics. For more information and data, see *American Demographics* (Vol. 17, No. 6).

CHAPTER REVIEW PROBLEMS

For each hypothesis test in Problems 1–8, please provide the following information.

(i) What is the level of significance? State the null and alternate hypotheses.
(ii) What sampling distribution will you use? What assumptions are you making? What is the value of the sample test statistic?
(iii) Find (or estimate) the *P*-value. Sketch the sampling distribution and show the area corresponding to the *P*-value.
(iv) Based on your answers in parts (i) to (iii), will you reject or fail to reject the null hypothesis? Are the data statistically significant at level α?
(v) State your conclusion in the context of the application.

Note: For degrees of freedom *d.f.* not in the Student's *t* table, use the closest *d.f.* that is smaller. In some cases this choice will increase the *P*-value by a small amount or increase the length of a confidence interval, thereby making the answer slightly more "conservative." Answers may vary due to rounding.

1. *Agriculture: Bell Peppers* The following data represent soil water content (percent water by volume) for independent random samples of soil taken from two experimental fields growing bell peppers (Reference: *Journal of Agricultural, Biological, and Environmental Statistics*). *Note:* These data are also available with other software on the statSpace CD-ROM.

Soil water content from field I: x_1; $n_1 = 72$

15.1	11.2	10.3	10.8	16.6	8.3	9.1	12.3	9.1	14.3
10.7	16.1	10.2	15.2	8.9	9.5	9.6	11.3	14.0	11.3
15.6	11.2	13.8	9.0	8.4	8.2	12.0	13.9	11.6	16.0
9.6	11.4	8.4	8.0	14.1	10.9	13.2	13.8	14.6	10.2
11.5	13.1	14.7	12.5	10.2	11.8	11.0	12.7	10.3	10.8
11.0	12.6	10.8	9.6	11.5	10.6	11.7	10.1	9.7	9.7
11.2	9.8	10.3	11.9	9.7	11.3	10.4	12.0	11.0	10.7
8.8	11.1								

Soil water content from field II: x_2; $n_2 = 80$

12.1	10.2	13.6	8.1	13.5	7.8	11.8	7.7	8.1	9.2
14.1	8.9	13.9	7.5	12.6	7.3	14.9	12.2	7.6	8.9
13.9	8.4	13.4	7.1	12.4	7.6	9.9	26.0	7.3	7.4
14.3	8.4	13.2	7.3	11.3	7.5	9.7	12.3	6.9	7.6
13.8	7.5	13.3	8.0	11.3	6.8	7.4	11.7	11.8	7.7
12.6	7.7	13.2	13.9	10.4	12.8	7.6	10.7	10.7	10.9
12.5	11.3	10.7	13.2	8.9	12.9	7.7	9.7	9.7	11.4
11.9	13.4	9.2	13.4	8.8	11.9	7.1	8.5	14.0	14.2

(a) Use a calculator with mean and standard deviation keys to verify that $\bar{x}_1 \approx 11.42$, $s_1 \approx 2.08$, $\bar{x}_2 \approx 10.65$, and $s_2 \approx 3.03$.

(b) Let μ_1 be the population mean for x_1 and let μ_2 be the population mean for x_2. Find a 95% confidence interval for $\mu_1 - \mu_2$.

(c) Examine the confidence interval and explain what it means in the context of this problem. Does the interval consist of numbers that are all positive? all negative? of different signs? At the 95% level of confidence, is the population mean soil water content of the first field higher than that of the second field?

(d) Which distribution (standard normal or Student's t) did you use? Why? Do you need information about the soil water content distributions?

(e) Use $\alpha = 0.01$ to test the claim that the population mean soil water content of the first field is higher than that of the second.

2. *Stocks: Retail and Utility* How profitable are different sectors of the stock market? One way to answer such a question is to examine profit as a percentage of stockholder equity. A random sample of 32 retail stocks such as Toys 'Я' Us, Best Buy, and Gap was studied for x_1, profit as a percentage of stockholder equity. The result was $\bar{x}_1 = 13.7$. A random sample of 34 utility (gas and electric) stocks such as Boston Edison, Wisconsin Energy, and Texas Utilities was studied for x_2, profit as a percentage of stockholder equity. The result was $\bar{x}_2 = 10.1$ (Source: *Fortune 500*, Vol. 135, No. 8). Assume $\sigma_1 = 4.1$ and $\sigma_2 = 2.7$.

(a) Let μ_1 represent the population mean profit as a percentage of stockholder equity for retail stocks, and let μ_2 represent the population mean profit as a percentage of stockholder equity for utility stocks. Find a 95% confidence interval for $\mu_1 - \mu_2$.

(b) Examine the confidence interval and explain what it means in the context of this problem. Does the interval consist of numbers that are all positive? all negative? of different signs? At the 95% level of confidence, does it appear that the profit as a percentage of stockholder equity for retail stocks is higher than that for utility stocks?

(c) Test the claim that the profit as a percentage of stockholder equity for retail stocks is higher than that for utility stocks. Use $\alpha = 0.01$.

3. *Wildlife: Wolves* A random sample of 18 adult male wolves from the Canadian Northwest Territories gave an average weight $\bar{x}_1 = 98$ lb with estimated sample standard deviation $s_1 = 6.5$ lb. Another sample of 24 adult male wolves from Alaska gave an average weight $\bar{x}_2 = 90$ lb with estimated sample standard deviation $s_2 = 7.3$ lb (Source: *The Wolf*, by L. D. Mech, University of Minnesota Press).

(a) Let μ_1 represent the population mean weight of adult male wolves from the Northwest Territories, and let μ_2 represent the population mean weight of adult male wolves from Alaska. Find a 75% confidence interval for $\mu_1 - \mu_2$.

(b) Examine the confidence interval and explain what it means in the context of this problem. Does the interval consist of numbers that are all positive? all negative? of different signs? At the 75% level of confidence, does it appear that the average weight of adult male wolves from the Northwest Territories is greater than that of the Alaska wolves?

(c) Test the claim that the average weight of adult male wolves from the Northwest Territories is different from that of Alaska wolves. Use $\alpha = 0.01$.

4. *Wildlife: Wolves* A random sample of 17 wolf litters in Ontario, Canada, gave an average of $\bar{x}_1 = 4.9$ wolf pups per litter with estimated sample standard deviation $s_1 = 1.0$. Another random sample of 6 wolf litters in Finland gave an average of $\bar{x}_2 = 2.8$ wolf pups per litter with sample standard deviation $s_2 = 1.2$ (see source for Problem 3).
   (a) Find an 85% confidence interval for $\mu_1 - \mu_2$, the difference in population mean litter size between Ontario and Finland.
   (b) Examine the confidence interval and explain what it means in the context of this problem. Does the interval consist of numbers that are all positive? all negative? of different signs? At the 85% level of confidence, does it appear that the average litter size of wolf pups in Ontario is greater than the average litter size in Finland?
   (c) Test the claim that the average litter size of wolf pups in Ontario is greater than the average litter size of wolf pups in Finland. Use $\alpha = 0.01$.

5. *Survey Response: Validity* The book *Survey Responses: An Evaluation of Their Validity*, by E. J. Wentland and K. Smith (Academic Press), includes studies reporting accuracy of answers to questions from surveys. A study by Locander et al. considered the question, "Are you a registered voter?" Accuracy of response was confirmed by a check of city voting records. Two methods of survey were used: a face-to-face interview and a telephone interview. A random sample of 93 people was asked the voter registration question face to face. Seventy-nine respondents gave accurate answers (as verified by city records). Another random sample of 83 people was asked the same question during a telephone interview. Seventy-four respondents gave accurate answers. Assume that the samples are representative of the general population.
   (a) Let p_1 be the population proportion of all people who answer the voter registration question accurately during a face-to-face interview. Let p_2 be the population proportion of all people who answer the question accurately during a telephone interview. Find a 95% confidence interval for $p_1 - p_2$.
   (b) Does the interval contain numbers that are all positive? all negative? mixed? Comment on the meaning of the confidence interval in the context of this problem. At the 95% level, do you detect any difference in the proportion of accurate responses from face-to-face interviews compared with the proportion of accurate responses from telephone interviews?
   (c) Test the claim that there is a difference in the proportion of accurate responses from face-to-face interviews compared with telephone interviews. Use $\alpha = 0.05$.

6. *Survey Response: Validity* Locander et al. (see reference in Problem 5) also studied the accuracy of responses on questions involving more sensitive material than voter registration. From public records, individuals were identified as having been charged with drunken driving not less than 6 months or more than 12 months from the starting date of the study. Two random samples from this group were studied. In the first sample of 30 individuals, the respondents were asked in a face-to-face interview if they had been charged with drunken driving in the last 12 months. Of these 30 people interviewed face-to-face, 16 answered the question accurately. The second random sample consisted of 46 people who had been charged with drunken driving. During a telephone interview, 25 of these responded accurately to the question asking if they had been charged with drunken driving during the past 12 months. Assume that the samples are representative of all people recently charged with drunken driving.

(a) Let p_1 represent the population proportion of all people with recent charges of drunken driving who respond accurately to a face-to-face interview asking if they have been charged with drunken driving during the past 12 months. Let p_2 represent the population proportion of people who respond accurately to the same question when it is asked in a telephone interview. Find a 90% confidence interval for $p_1 - p_2$.

(b) Does the interval found in part (a) contain numbers that are all positive? all negative? mixed? Comment on the meaning of the confidence interval in the context of this problem. At the 90% level, do you detect any differences in the proportion of accurate responses to the question from face-to-face interviews as compared with the proportion of accurate responses from telephone interviews?

(c) Test the claim that there is a difference in the proportion of accurate responses from face-to-face interviews compared with the proportion of accurate responses from telephone interviews. Use $\alpha = 0.05$.

7. *Marketing: Sporting Goods* A marketing consultant was hired to visit a random sample of five sporting goods stores across the state of California. Each store was part of a large franchise of sporting goods stores. The consultant taught the managers of each store better ways to advertise and display their goods. The net sales for 1 month before and 1 month after the consultant's visit were recorded as follows for each store (in thousands of dollars):

Store	1	2	3	4	5
Before visit	57.1	94.6	49.2	77.4	43.2
After visit	63.5	101.8	57.8	81.2	41.9

Do the data indicate that the average net sales improved? (Use $\alpha = 0.05$.)

8. *Psychology: Creative Thinking* Six sets of identical twins were randomly selected from a population of identical twins. One child was taken at random from each pair to form an experimental group. These children participated in a program designed to promote creative thinking. The other child from each pair was part of the control group that did not participate in the program to promote creative thinking. At the end of the program, a creative problem-solving test was given with the results shown in the following table:

Twin pair	A	B	C	D	E	F
Experimental group	53	35	12	25	33	47
Control group	39	21	5	18	21	42

Higher scores indicate better performance in creative problem solving. Do the data support the claim that the program of the experimental group did promote creative problem solving? (Use $\alpha = 0.01$.)

DATA HIGHLIGHTS: GROUP PROJECTS

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

"Sweets May Not Be Culprit in Hyper Kids" was a *USA Today* (February 3, 1994) headline reporting results of a study that appeared in the *New England Journal of Medicine.*

In this study, the subjects were 25 normal preschoolers aged 3 to 5, and 23 kids aged 6 to 10, who had been described as "sensitive to sugar." The kids and their families were put on three different diets for 3 weeks each. One diet was high in sugar, one was low in sugar and contained aspartame, and one was low in sugar and contained saccharin. The diets were all free of additives, artificial food coloring, preservatives, and chocolate. All food in the households was removed, and then meals were delivered to the families. Researchers gathered information about the kids' behavior from parents, babysitters, and teachers. In addition, researchers tested the kids for memory, concentration, reading, and math skills. The result: "We couldn't find any difference in terms of their behavior or their learning on any of the three diets," said Mark Wolraich, professor of pediatrics at Vanderbilt University Medical Center, who oversaw the project. In another interview, Dr. Wolraich was quoted as saying, "Our study would say there is no evidence sugar has an adverse effect on children's behavior."

(a) This research involved comparisons of several means, not just two. However, let us take a simplified view of the problem and consider the difference in behavior when children consumed the diet with sugar compared with their behavior when they consumed the diet with aspartame and low sugar. List some variables that might be measured to reflect the behavior of the children.

(b) Let's assume that the general null hypothesis was that there is no difference in children's behavior when they have a diet high in sugar. Was the evidence sufficient to allow the researchers to reject the null hypothesis and conclude that there are differences in children's behavior when they have a diet high in sugar? When we cannot reject H_0, have we *proved* that H_0 is true? In your own words, paraphrase the comments made by Dr. Wolraich.

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.

Is there a relationship between confidence intervals and two-tailed hypothesis tests? The answer is yes. Let c be the level of confidence used to construct a confidence interval from sample data. Let α be the level of significance for a two-tailed hypothesis test. The following statement applies to hypothesis tests of the mean.

> For a two-tailed hypothesis test with level of significance α and null hypothesis H_0: $\mu = k$, we *reject* H_0 whenever k falls *outside* the $c = 1 - \alpha$ confidence interval for μ based on the sample data. When k falls within the $c = 1 - \alpha$ confidence interval, we do not reject H_0.
>
> For a one-tailed hypothesis test with level of significance α and null hypothesis H_0: $\mu = k$, we *reject* H_0 whenever k falls *outside* the $c = 1 - 2\alpha$ confidence interval for μ based on the sample data. When k falls within the $c = 1 - 2\alpha$ confidence interval, we do not reject H_0.

A corresponding relationship between confidence intervals and two-tailed hypothesis tests is also valid for other parameters such as p, $\mu_1 - \mu_2$, and $p_1 - p_2$.

(a) Consider the hypotheses H_0: $\mu_1 - \mu_2 = 0$ and H_1: $\mu_1 - \mu_2 \neq 0$. Suppose a 95% confidence interval for $\mu_1 - \mu_2$ contains only positive numbers. Should you reject the null hypothesis when $\alpha = 0.05$? Why or why not?

(b) Consider the hypotheses H_0: $p_1 - p_2 = 0$ and H_1: $p_1 - p_2 > 0$. Suppose a 98% confidence interval for $p_1 - p_2$ contains only positive numbers. Should you reject or fail to reject H_0 at the $\alpha = 0.01$ level of significance?

Using Technology

TI-84PLUS/TI-83PLUS • EXCEL • MINITAB • SPSS

APPLICATION

Paired Difference Test

Suppose a random sample of eight nurses' schedules were changed from the day shift to a rotating shift of some night work and some day work. For each of these nurses, the information shown in the table below was recorded about level of tension and anxiety at the end of a day shift and also at the end of a night shift.

Data for Nurses Working Both Shifts

Nurse	1	2	3	4	5	6	7	8
Day shift (B)	1.5	3	2	3	2	2	1	2
Night shift (A)	3.5	2	4	4	3.5	2	3.5	3

(a) Explain why the sample data for the day shift cannot be thought of as independent of the sample data for the night shift.

(b) Let us say that A is the random variable representing tension levels of night nurses and B is the random variable representing tension levels of day nurses. If we wanted to test the claim that nurses have a higher level of tension after the night shift, what would we use for the null hypothesis? What would we use for the alternate hypothesis? Choose the appropriate hypotheses and enter your choices on the computer. Use a 2% level of significance. Shall we accept or reject the claim that after a night shift, nurses express more feelings of tension on average than they do after a day shift?

(c) What is the smallest level of significance at which these data will allow us to accept the claim that after a night shift, nurses express more feelings of tension?

Technology Hints: Paired Difference Test

TI-84Plus/TI-83Plus

Enter the first number of each data pair in the list L_1 and the corresponding value in the second list L_2. Create the list L_3 by subtracting L_2 from L_1. List L_3 contains the differences d. Our next step is to conduct a t test on the differences in L_3. Press **STAT**, select **TESTS**, and choose **Option 2:T-Test.** Use **Inpt:Data.** The value for μ_0 is 0 because we are testing the hypothesis H_0: $\mu_0 = 0$. Indicate that the data are in list L_3 with **Freq: 1.** Select the appropriate symbol ($\neq$, $<$, or $>$) for the alternate hypothesis, and then choose **Calculate,** which gives the value of the sample test statistic along with its t value and P-value, or **Draw,** which shows the area corresponding to the P-value of the sample test statistic and gives the t value and P-value of the sample test statistic.

Excel

Enter the data in two columns. Then use the menu choices ➤ **Tools** ➤ **Data Analysis** ➤ ***t*-Test: Paired Two-Sample for Means.** The output contains the mean and variance for each data column, the degrees of freedom for the test, the sample t statistic, the P-value for a one-tailed test, the P-value for a two-tailed test, and the critical values t_0 for the specified level of significance for both a one-tailed test and a two-tailed test.

Minitab

Enter the data in two columns. Then use the menu selections ➤ **Stats** ➤ **Basic Statistics** ➤ **Paired-*t*.** The output includes an option for a confidence interval. Other items include the number of data pairs n, the mean, the standard deviation, and the standard error of the mean, for each data set as well as for the differences. The t value of the sample test statistic is given along with the P-value. There is also an option to generate a histogram, dotplot, or boxplot for the differences.

SPSS

Enter the data in two columns in the Data View window. In the Variable View window, name and label the two columns. Be sure the data type is numeric. Use the commands **Analyze ➤ Compare Means ➤ Paired-Samples T Test.** In the dialogue box, select Options to set the confidence level for a confidence interval. Results for a two-tailed test with H_0: $\mu_d = 0$ and H_1: $\mu_d \neq 0$ are given, with test statistic, standard deviation of the paired differences, standard error, t value of the test statistic, degrees of freedom, and P-value of the test statistic. In the output, **Sig(2-tailed)** is the P-value of the test statistic for a two-tailed test. A confidence interval for the mean of the paired differences is also given, as well as the mean, standard deviation, and standard error for each column of data.

OTHER APPLICATIONS

Technology Hints: Inferences for $\mu_1 - \mu_2$

TI-84Plus/TI-83Plus

You have the option of using raw data entered into two separate lists or summary statistics. Press the **STAT** key, select the **TESTS** option, and then use option **3:2-SampZTest** or option **4:2-SampTTest.** The output gives the z or t value of the sample test statistic and the P-value for the test. For confidence intervals, use option **9:2-SampZInt** or option **0:2-SampTTest** under the Tests option.

Excel

Enter the data in two columns. Then use the menu selection **➤ Tools ➤ Data Analysis ➤ z-Test Two Sample for Means** if you know σ_1 and σ_2. Otherwise, use **t-Test:Two Sample Assuming Unequal Variances.** The output provides the mean and variance for each variable, the z or t value of the sample test statistic, the P-values for a one-tailed test and a two-tailed test, and the critical z_0 or t_0 values for a one-tailed test and a two-tailed test.

Minitab

When testing the difference of means from independent samples, Minitab always uses the t distribution. The P-value of the sample test statistic will be slightly larger than the P-value generated using the normal distribution. Enter the data in two columns. Then use the menu selections **➤ Stat ➤ Basic Statistics ➤ 2-Sample t.** Do not check equal variances. The output gives an option for a confidence interval as well as for dotplots or boxplots of the two variables. The output displays the mean, standard deviation, and standard error of the mean for each of the variables. The t value of the sample test statistic and the P-value are based on the Student's t distribution.

SPSS

Enter all the data in one column. In an adjacent column, enter a grouping variable with two values to separate the cases into two independent samples. For instance, use the value 1 to indicate that a data value is from the first sample, and the value 2 to indicate that a data value is from the second sample. Then use the menu choices **Analyze ➤ Compare Means ➤ Independent-Samples T Test.** In the dialogue box, move the column containing the data into the Test Variable box and the column containing the grouping variables into the Grouping Variable box. Then press Define Groups and enter the grouping values, such as 1 and 2. A confidence interval can be set by using the Options button. The output gives the mean, standard deviation, and standard error for each sample. Test results show the t value, degrees of freedom, and P-value for a two-tailed test for both the case of equal variances and the case of unequal variances. Confidence intervals are also included in the output.