Solution

Step 1 State the hypothesis and identify the claim. Since we are interested to see if there has been an increase in deposits, the deposits 3 years ago must be less than the deposits today; hence, the differences must be significantly less 3 years ago than they are today. Hence the mean of the differences must be less than zero.

H_0: $\mu_D = 0$ and H_1: $\mu_D < 0$ (claim)

Step 2 Find the critical value. The degrees of freedom are $n - 1$, or $9 - 1 = 8$. The critical value for a left-tailed test with $\alpha = 0.05$ is -1.860.

Step 3 Compute the test value.

a. Make a table.

3 years ago (X_1)	Now (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
11.42	16.69		
8.41	9.44		
3.98	6.53		
7.37	5.58		
2.28	2.92		
1.10	1.88		
1.00	1.78		
0.90	1.50		
1.35	1.22		

b. Find the differences and place the results in column A.

$11.42 - 16.69 = -5.27$
$8.41 - 9.44 = -1.03$
$3.98 - 6.53 = -2.55$
$7.37 - 5.58 = +1.79$
$2.28 - 2.92 = -0.64$
$1.10 - 1.88 = -0.78$
$1.00 - 1.78 = -0.78$
$0.9 - 1.50 = -0.60$
$1.35 - 1.22 = +0.13$
$\Sigma D = 9.73$

c. Find the means of the differences.

$$\overline{D} = \frac{\Sigma D}{n} = \frac{-9.73}{9} = -1.081$$

d. Square the differences and place the results in Column B.

$(-5.27)^2 = 27.7729$
$(-1.03)^2 = 1.0609$
$(-2.55)^2 = 6.5025$
$(+1.79)^2 = 3.2041$
$(-0.64)^2 = 0.4096$
$(-0.78)^2 = 0.6084$
$(-0.78)^2 = 0.6084$
$(-0.60)^2 = 0.3600$
$(+0.13)^2 = 0.1690$
$\Sigma D^2 = 40.5437$

The completed table is shown next.

3 years ago (X_1)	Now (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
11.42	16.69	−5.27	27.7299
8.41	9.44	−1.03	1.0609
3.98	6.53	−2.55	6.5025
7.37	5.58	+1.79	3.2041
2.28	2.92	−0.64	0.4096
1.10	1.88	−0.78	0.6084
1.00	1.78	−0.78	0.6084
0.90	1.58	−0.60	0.3600
1.35	1.22	+0.13	0.1690
		$\Sigma D = 9.73$	$\Sigma D^2 = 40.5437$

e. Find the standard deviation of the differences.

$$s_D = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$$

$$= \sqrt{\frac{9(40.5437) - (9.73)^2}{9(9-1)}}$$

$$= \sqrt{\frac{270.2204}{72}}$$

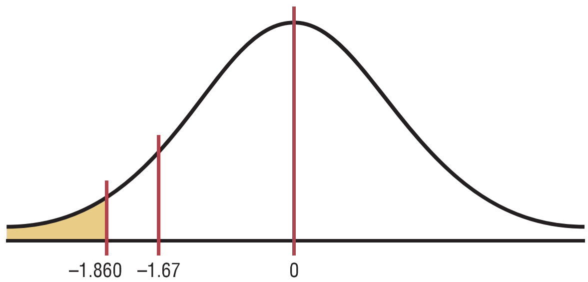
$$= 1.937$$

f. Find the test value.

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}} = \frac{-1.081 - 0}{1.937/\sqrt{9}} = -1.67$$

Step 4 Make the decision. Do not reject the null hypothesis since the test value, −1.67, is greater than the critical value, −1.860. See Figure 9–6.

Figure 9–6
Critical and Test Values for Example 9–6



Step 5 Summarize the results. There is not enough evidence to show that the deposits have increased over the last 3 years.

The formulas for this t test are summarized next.

Formulas for the t Test for Dependent Samples

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}}$$

with d.f. $= n - 1$ and where

$$\overline{D} = \frac{\Sigma D}{n} \qquad \text{and} \qquad s_D = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$$

Example 9–7

Cholesterol Levels

A dietitian wishes to see if a person's cholesterol level will change if the diet is supplemented by a certain mineral. Six subjects were pretested, and then they took the mineral supplement for a 6-week period. The results are shown in the table. (Cholesterol level is measured in milligrams per deciliter.) Can it be concluded that the cholesterol level has been changed at $\alpha = 0.10$? Assume the variable is approximately normally distributed.

Subject	1	2	3	4	5	6
Before (X_1)	210	235	208	190	172	244
After (X_2)	190	170	210	188	173	228

Solution

Step 1 State the hypotheses and identify the claim. If the diet is effective, the before cholesterol levels should be different from the after levels.

H_0: $\mu_D = 0$ and H_1: $\mu_D \neq 0$ (claim)

Step 2 Find the critical value. The degrees of freedom are 5. At $\alpha = 0.10$, the critical values are ± 2.015.

Step 3 Compute the test value.

a. Make a table.

Before (X_1)	**After (X_2)**	**A** $D = X_1 - X_2$	**B** $D^2 = (X_1 - X_2)^2$
210	190		
235	170		
208	210		
190	188		
172	173		
244	228		

b. Find the differences and place the results in column A.

$210 - 190 = 20$
$235 - 170 = 65$
$208 - 210 = -2$
$190 - 188 = 2$
$172 - 173 = -1$
$244 - 228 = 16$
$\Sigma D = 100$

c. Find the mean of the differences.

$$\overline{D} = \frac{\Sigma D}{n} = \frac{100}{6} = 16.7$$

d. Square the differences and place the results in column B.

$$\begin{aligned}(20)^2 &= 400\\(65)^2 &= 4225\\(-2)^2 &= 4\\(2)^2 &= 4\\(-1)^2 &= 1\\(16)^2 &= \underline{256}\\\Sigma D^2 &= 4890\end{aligned}$$

Then complete the table as shown.

Before (X_1)	After (X_2)	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
210	190	20	400
235	170	65	4225
208	210	−2	4
190	188	2	4
172	173	−1	1
244	228	16	256
		$\Sigma D = 100$	$\Sigma D^2 = 4890$

e. Find the standard deviation of the differences.

$$\begin{aligned}s_D &= \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}\\&= \sqrt{\frac{6 \cdot 4890 - 100^2}{6(6-1)}}\\&= \sqrt{\frac{29{,}340 - 10{,}000}{30}}\\&= 25.4\end{aligned}$$

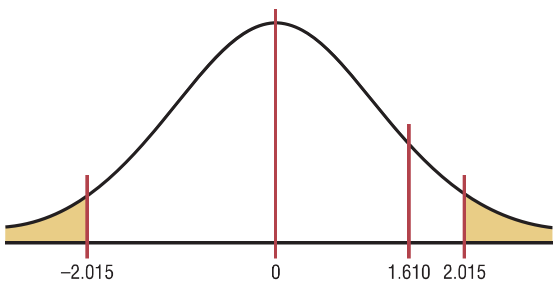
f. Find the test value.

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}} = \frac{16.7 - 0}{25.4/\sqrt{6}} = 1.610$$

Step 4 Make the decision. The decision is to not reject the null hypothesis, since the test value 1.610 is in the noncritical region, as shown in Figure 9–7.

Figure 9–7

Critical and Test Values for Example 9–7



Step 5 Summarize the results. There is not enough evidence to support the claim that the mineral changes a person's cholesterol level.

The steps for this t test are summarized in the Procedure Table.

Procedure Table

Testing the Difference Between Means for Dependent Samples

Step 1 State the hypotheses and identify the claim.

Step 2 Find the critical value(s).

Step 3 Compute the test value.

a. Make a table, as shown.

X_1	X_2	A $D = X_1 - X_2$	B $D^2 = (X_1 - X_2)^2$
⋮	⋮		
		$\Sigma D =$ ______	$\Sigma D^2 =$ ______

b. Find the differences and place the results in column A.

$$D = X_1 - X_2$$

c. Find the mean of the differences.

$$\overline{D} = \frac{\Sigma D}{n}$$

d. Square the differences and place the results in column B. Complete the table.

$$D^2 = (X_1 - X_2)^2$$

e. Find the standard deviation of the differences.

$$s_D = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$$

f. Find the test value.

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}} \qquad \text{with d.f.} = n - 1$$

Step 4 Make the decision.

Step 5 Summarize the results.

Unusual Stat

About 4% of Americans spend at least one night in jail each year.

The P-values for the t test are found in Table F. For a two-tailed test with d.f. $= 5$ and $t = 1.610$, the P-value is found between 1.476 and 2.015; hence, $0.10 < P\text{-value} < 0.20$. Thus, the null hypothesis cannot be rejected at $\alpha = 0.10$.

If a specific difference is hypothesized, this formula should be used

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}}$$

where μ_D is the hypothesized difference.

For example, if a dietitian claims that people on a specific diet will lose an average of 3 pounds in a week, the hypotheses are

$$H_0\colon \mu_D = 3 \qquad \text{and} \qquad H_1\colon \mu_D \neq 3$$

The value 3 will be substituted in the test statistic formula for μ_D.

Confidence intervals can be found for the mean differences with this formula.

Confidence Interval for the Mean Difference

$$\overline{D} - t_{\alpha/2}\frac{s_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\alpha/2}\frac{s_D}{\sqrt{n}}$$

d.f. $= n - 1$

Example 9–8

Find the 90% confidence interval for the data in Example 9–7.

Solution

Substitute in the formula.

$$\overline{D} - t_{\alpha/2}\frac{s_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\alpha/2}\frac{s_D}{\sqrt{n}}$$

$$16.7 - 2.015 \cdot \frac{25.4}{\sqrt{6}} < \mu_D < 16.7 + 2.015 \cdot \frac{25.4}{\sqrt{6}}$$

$$16.7 - 20.89 < \mu_D < 16.7 + 20.89$$

$$-4.19 < \mu_D < 37.59$$

Since 0 is contained in the interval, the decision is to not reject the null hypothesis $H_0\colon \mu_D = 0$.

Speaking of **Statistics**

Can Video Games Save Lives?

Can playing video games help doctors perform surgery? The answer is yes. A study showed that surgeons who played video games for at least 3 hours each week made about 37% fewer mistakes and finished operations 27% faster than those who did not play video games.

The type of surgery that they performed is called *laparoscopic* surgery, where the surgeon inserts a tiny video camera into the body and uses a joystick to maneuver the surgical instruments while watching the results on a television monitor. This study compares two groups and uses proportions. What statistical test do you think was used to compare the percentages? (See Section 9–4.)

Applying the Concepts 9–3

Air Quality

As a researcher for the EPA, you have been asked to determine if the air quality in the United States has changed over the past 2 years. You select a random sample of 10 metropolitan areas and find the number of days each year that the areas failed to meet acceptable air quality standards. The data are shown.

Year 1	18	125	9	22	138	29	1	19	17	31
Year 2	24	152	13	21	152	23	6	31	34	20

Source: *The World Almanac and Book of Facts.*

Based on the data, answer the following questions.

1. What is the purpose of the study?
2. Are the samples independent or dependent?
3. What hypotheses would you use?
4. What is (are) the critical value(s) that you would use?
5. What statistical test would you use?
6. How many degrees of freedom are there?
7. What is your conclusion?
8. Could an independent means test have been used?
9. Do you think this was a good way to answer the original question?

See page 531 for the answers.

Exercises 9–3

1. Classify each as independent or dependent samples.

a. Heights of identical twins Dependent

b. Test scores of the same students in English and psychology Dependent

c. The effectiveness of two different brands of aspirin Independent

d. Effects of a drug on reaction time, measured by a before-and-after test Dependent

e. The effectiveness of two different diets on two different groups of individuals Independent

For Exercises 2 through 10, perform each of these steps. Assume that all variables are normally or approximately normally distributed.

a. State the hypotheses and identify the claim.

b. Find the critical value(s).

c. Compute the test value.

d. Make the decision.

e. Summarize the results.

Use the traditional method of hypothesis testing unless otherwise specified.

2. Retention Test Scores A sample of non-English majors at a selected college was used in a study to see if the student retained more from reading a 19th-century novel or by watching it in DVD form. Each student was assigned one novel to read and a different one to watch, and then they were given a 20-point written quiz on each novel. The test results are shown below. At $\alpha = 0.05$, can it be concluded that the book scores are higher than the DVD scores?

Book	90	80	90	75	80	90	84
DVD	85	72	80	80	70	75	80

3. Improving Study Habits As an aid for improving students' study habits, nine students were randomly selected to attend a seminar on the importance of education in life. The table shows the number of hours each student studied per week before and after the seminar. At $\alpha = 0.10$, did attending the seminar increase the number of hours the students studied per week?

Before	9	12	6	15	3	18	10	13	7
After	9	17	9	20	2	21	15	22	6

4. Obstacle Course Times An obstacle course was set up on a campus, and 10 volunteers were given a chance to complete it while they were being timed. They then sampled a new energy drink and were given the opportunity to run the course again. The "before" and "after" times in seconds are shown below. Is there sufficient evidence at $\alpha = 0.05$ to conclude that the students did better the second time? Discuss possible reasons for your results.

Student	1	2	3	4	5	6	7	8
Before	67	72	80	70	78	82	69	75
After	68	70	76	65	75	78	65	68

5. Sleep Report Students in a statistics class were asked to report the number of hours they slept on weeknights and on weekends. At $\alpha = 0.05$, is there sufficient evidence that there is a difference in the mean number of hours slept?

Student	1	2	3	4	5	6	7	8
Hours, Sun.–Thurs.	8	5.5	7.5	8	7	6	6	8
Hours, Fri.–Sat.	4	7	10.5	12	11	9	6	9

6. PGA Golf Scores At a recent PGA tournament (the Honda Classic at Palm Beach Gardens, Florida) the following scores were posted for eight randomly selected golfers for two consecutive days. At $\alpha = 0.05$, is there evidence of a difference in mean scores for the two days?

Golfer	1	2	3	4	5	6	7	8
Thursday	67	65	68	68	68	70	69	70
Friday	68	70	69	71	72	69	70	70

Source: *Washington Observer-Reporter.*

7. Reducing Errors in Grammar A composition teacher wishes to see whether a new grammar program will reduce the number of grammatical errors her students make when writing a two-page essay. The data are shown here. At $\alpha = 0.025$, can it be concluded that the number of errors has been reduced?

Student	1	2	3	4	5	6
Errors before	12	9	0	5	4	3
Errors after	9	6	1	3	2	3

8. Overweight Dogs A veterinary nutritionist developed a diet for overweight dogs. The total volume of food consumed remains the same, but one-half of the dog food is replaced with a low-calorie "filler" such as canned green beans. Six overweight dogs were randomly selected from her practice and were put on this program. Their initial weights were recorded, and then they were weighed again after 4 weeks. At the 0.05 level of significance can it be concluded that the dogs lost weight?

Before	42	53	48	65	40	52
After	39	45	40	58	42	47

9. Pulse Rates of Identical Twins A researcher wanted to compare the pulse rates of identical twins to see whether there was any difference. Eight sets of twins were selected. The rates are given in the table as number of beats per minute. At $\alpha = 0.01$, is there a significant difference in the average pulse rates of twins? Find the 99% confidence interval for the difference of the two. Use the P-value method.

Twin A	87	92	78	83	88	90	84	93
Twin B	83	95	79	83	86	93	80	86

10. A random sample of six music students played a short song, and the number of mistakes each student made was recorded. After they practiced the song 5 times, the number of mistakes each student made was recorded. The data are shown. At $\alpha = 0.05$, can it be concluded that there was a decrease in the mean number of mistakes?

Student	A	B	C	D	E	F
Before	10	6	8	8	13	8
After	4	2	2	7	8	9

Extending the Concepts

11. Instead of finding the mean of the differences between X_1 and X_2 by subtracting $X_1 - X_2$, you can find it by finding the means of X_1 and X_2 and then subtracting the means. Show that these two procedures will yield the same results.

Technology *Step by Step*

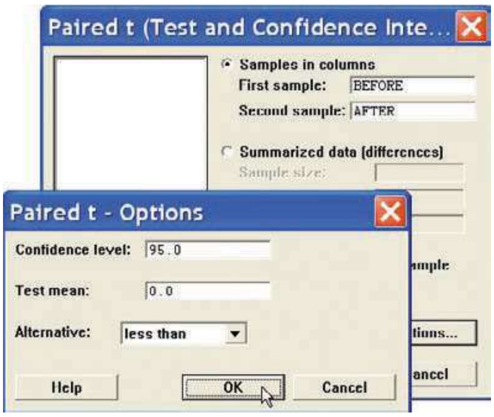
MINITAB
Step by Step

Test the Difference Between Two Means: Dependent Samples

A physical education director claims by taking a special vitamin, a weight lifter can increase his strength. Eight athletes are selected and given a test of strength, using the standard bench press. After 2 weeks of regular training, supplemented with the vitamin, they are tested again. Test the effectiveness of the vitamin regimen at $\alpha = 0.05$. Each value in these data represents the maximum number of pounds the athlete can bench-press. Assume that the variable is approximately normally distributed.

Athlete	1	2	3	4	5	6	7	8
Before (X_1)	210	230	182	205	262	253	219	216
After (X_2)	219	236	179	204	270	250	222	216

1. Enter the data into C1 and C2. Name the columns **Before** and **After.**
2. Select **Stat>Basic Statistics>Paired t.**
3. Double-click C1 Before for First sample.
4. Double-click C2 After for Second sample. The second sample will be subtracted from the first. The differences are not stored or displayed.
5. Click [Options].
6. Change the Alternative to less than.
7. Click [OK] twice.



Paired t-Test and CI: BEFORE, AFTER

```
Paired t for BEFORE - AFTER
            N      Mean      StDev    SE Mean
BEFORE      8   222.125     25.920      9.164
AFTER       8   224.500     27.908      9.867
Difference  8   -2.37500   4.83846    1.71065

95% upper bound for mean difference: 0.86597
t-Test of mean difference = 0 (vs < 0) : t-Value = -1.39  P-Value = 0.104.
```

Since the *P*-value is 0.104, do not reject the null hypothesis. The sample difference of −2.38 in the strength measurement is not statistically significant.

TI-83 Plus or TI-84 Plus
Step by Step

Hypothesis Test for the Difference Between Two Means: Dependent Samples

1. Enter the data values into L_1 and L_2.
2. Move the cursor to the top of the L_3 column so that L_3 is highlighted.
3. Type **$L_1 - L_2$**, then press **ENTER.**
4. Press **STAT** and move the cursor to TESTS.
5. Press **2** for TTest.
6. Move the cursor to Data and press **ENTER.**

7. Type in the appropriate values, using **0** for μ_0 and $\mathbf{L_3}$ for the list.
8. Move the cursor to the appropriate alternative hypothesis and press **ENTER.**
9. Move the cursor to Calculate and press **ENTER.**

Confidence Interval for the Difference Between Two Means: Dependent Samples

1. Enter the data values into L_1 and L_2.
2. Move the cursor to the top of the L_3 column so that L_3 is highlighted.
3. Type $\mathbf{L_1 - L_2}$, then press **ENTER.**
4. Press **STAT** and move the cursor to TESTS.
5. Press **8** for TInterval.
6. Move the cursor to Stats and press **ENTER.**
7. Type in the appropriate values, using $\mathbf{L_3}$ for the list.
8. Move the cursor to Calculate and press **ENTER.**

Excel
Step by Step

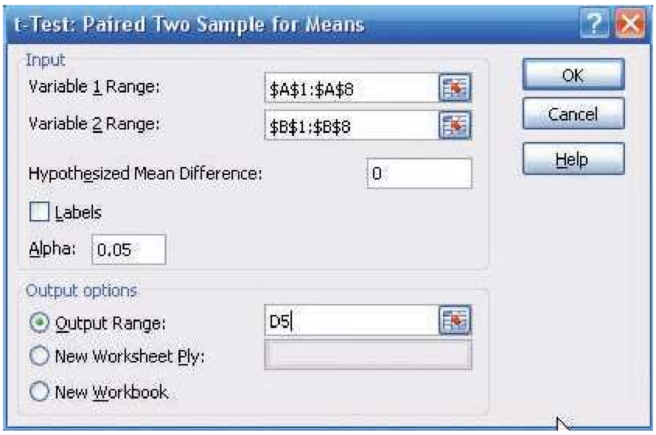
Testing the Difference Between Two Means: Dependent Samples

Example XL9–3

Test the claim that there is no difference between population means based on these sample paired data. Use $\alpha = 0.05$.

Set A	33	35	28	29	32	34	30	34
Set B	27	29	36	34	30	29	28	24

1. Enter the 8-number data set A into column A.
2. Enter the 8-number data set B into column B.
3. Select the Data tab from the toolbar. Then select Data Analysis.
4. In the Data Analysis box, under Analysis Tools select *t*-test: Paired Two Sample for Means, and click [OK].
5. In Input, type in the Variable 1 Range: **A1:A8** and the Variable 2 Range: **B1:B8.**
6. Type **0** for the Hypothesized Mean Difference.
7. Type **0.05** for Alpha.
8. In Output options, type **D5** for the Output Range, then click [OK].



t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	31.875	29.625
Variance	6.696428571	14.55357143
Observations	8	8
Pearson Correlation	-0.757913399	
Hypothesized Mean Difference	0	
df	7	
t Stat	1.057517468	
P(T<=t) one-tail	0.1626994	
t Critical one-tail	1.894578604	
P(T<=t) two-tail	0.3253988	
t Critical two-tail	2.364624251	

Note: You may need to increase the column width to see all the results. To do this:

1. Highlight the columns D, E, and F.
2. Select **Format>AutoFit** Column Width.

The output shows a *P*-value of 0.3253988 for the two-tailed case. This value is greater than the alpha level of 0.05, so we fail to reject the null hypothesis.

9–4 Testing the Difference Between Proportions

Objective 4

Test the difference between two proportions.

The z test with some modifications can be used to test the equality of two proportions. For example, a researcher might ask, Is the proportion of men who exercise regularly less than the proportion of women who exercise regularly? Is there a difference in the percentage of students who own a personal computer and the percentage of nonstudents who own one? Is there a difference in the proportion of college graduates who pay cash for purchases and the proportion of non-college graduates who pay cash?

Recall from Chapter 7 that the symbol $\hat{p}$ ("p hat") is the sample proportion used to estimate the population proportion, denoted by p. For example, if in a sample of 30 college students, 9 are on probation, then the sample proportion is $\hat{p} = \frac{9}{30}$, or 0.3. The population proportion p is the number of all students who are on probation, divided by the number of students who attend the college. The formula for $\hat{p}$ is

$$\hat{p} = \frac{X}{n}$$

where

X = number of units that possess the characteristic of interest
n = sample size

When you are testing the difference between two population proportions p_1 and p_2, the hypotheses can be stated thus, if no difference between the proportions is hypothesized.

$$H_0: p_1 = p_2 \qquad \text{or} \qquad H_0: p_1 - p_2 = 0$$
$$H_1: p_1 \neq p_2 \qquad \qquad \quad H_1: p_1 - p_2 \neq 0$$

Similar statements using $<$ or $>$ in the alternate hypothesis can be formed for one-tailed tests.

For two proportions, $\hat{p}_1 = X_1/n_1$ is used to estimate p_1 and $\hat{p}_2 = X_2/n_2$ is used to estimate p_2. The standard error of the difference is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\sigma^2_{p_1} + \sigma^2_{p_2}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

where $\sigma^2_{p_1}$ and $\sigma^2_{p_2}$ are the variances of the proportions, $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, and n_1 and n_2 are the respective sample sizes.

Since p_1 and p_2 are unknown, a weighted estimate of p can be computed by using the formula

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

and $\bar{q} = 1 - \bar{p}$. This weighted estimate is based on the hypothesis that $p_1 = p_2$. Hence, $\bar{p}$ is a better estimate than either $\hat{p}_1$ or $\hat{p}_2$, since it is a combined average using both $\hat{p}_1$ and $\hat{p}_2$.

Since $\hat{p}_1 = X_1/n_1$ and $\hat{p}_2 = X_2/n_2$, $\bar{p}$ can be simplified to

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Finally, the standard error of the difference in terms of the weighted estimate is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

The formula for the test value is shown next.

Formula for the *z* Test for Comparing Two Proportions

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\,\bar{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

where

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} \qquad \hat{p}_1 = \frac{X_1}{n_1}$$

$$\bar{q} = 1 - \bar{p} \qquad \hat{p}_2 = \frac{X_2}{n_2}$$

This formula follows the format

$$\text{Test value} = \frac{(\text{observed value}) - (\text{expected value})}{\text{standard error}}$$

Assumptions for the *z* Test for Two Proportions

1. The samples must be random samples.
2. The sample data are independent of one another.
3. For both samples $np \geq 5$ and $nq \geq 5$.

Example 9–9

Vaccination Rates in Nursing Homes

In the nursing home study mentioned in the chapter-opening Statistics Today, the researchers found that 12 out of 34 small nursing homes had a resident vaccination rate of less than 80%, while 17 out of 24 large nursing homes had a vaccination rate