22. Sale Prices for Houses The average sales price of new one-family houses in the Midwest is $250,000 and in the South is $253,400. A random sample of 40 houses in each region was examined with the following results. At the 0.05 level of significance can it be concluded that the difference in mean sales price for the two regions is greater than $3400?

	South	Midwest
Sample size	40	40
Sample mean	261,500	248,200
Population standard deviation	10,500	12,000

Source: *New York Times Almanac.*

23. Average Earnings for College Graduates The average earnings of year-round full-time workers with bachelor's degrees or more is $88,641 for men and $58,000 for women—a difference of slightly over $30,000 a year. One hundred of each were sampled, resulting in a sample mean of $90,200 for men, and the population standard deviation is $15,000, and a mean of $57,800 for women, and the population standard deviation is $12,800. At the 0.01 level of significance can it be concluded that the difference in means is not $30,000?

Source: *New York Times Almanac.*

Technology *Step by Step*

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

Hypothesis Test for the Difference Between Two Means and *z* Distribution (Data)

1. Enter the data values into L_1 and L_2.
2. Press **STAT** and move the cursor to TESTS.
3. Press **3** for 2-SampZTest.
4. Move the cursor to Data and press **ENTER.**
5. Type in the appropriate values.
6. Move the cursor to the appropriate alternative hypothesis and press **ENTER.**
7. Move the cursor to Calculate and press **ENTER.**

Hypothesis Test for the Difference Between Two Means and *z* Distribution (Statistics)

1. Press **STAT** and move the cursor to TESTS.
2. Press **3** for 2-SampZTest.
3. Move the cursor to Stats and press **ENTER.**
4. Type in the appropriate values.
5. Move the cursor to the appropriate alternative hypothesis and press **ENTER.**
6. Move the cursor to Calculate and press **ENTER.**

Confidence Interval for the Difference Between Two Means and *z* Distribution (Data)

1. Enter the data values into L_1 and L_2.
2. Press **STAT** and move the cursor to TESTS.
3. Press **9** for 2-SampZInt.
4. Move the cursor to Data and press **ENTER.**
5. Type in the appropriate values.
6. Move the cursor to Calculate and press **ENTER.**

Confidence Interval for the Difference Between Two Means and *z* Distribution (Statistics)

1. Press **STAT** and move the cursor to TESTS.
2. Press **9** for 2-SampZInt.
3. Move the cursor to Stats and press **ENTER.**
4. Type in the appropriate values.
5. Move the cursor to Calculate and press **ENTER.**

z-Test: Two Sample for Means		
	Variable 1	*Variable 2*
Mean	14.06666667	9.266666667
Known Variance	10.067	7.067
Observations	15	15
Hypothesized Mean Difference	0	
z	4.491149228	
P(Z<=z) one-tail	3.54522E-06	
z Critical one-tail	1.644853	
P(Z<=z) two-tail	7.09045E-06	
z Critical two-tail	1.959961082	

9–2 Testing the Difference Between Two Means of Independent Samples: Using the *t* Test

Objective 2

Test the difference between two means for independent samples, using the *t* test.

In Section 9–1, the *z* test was used to test the difference between two means when the population standard deviations were known and the variables were normally or approximately normally distributed, or when both sample sizes were greater than or equal to 30. In many situations, however, these conditions cannot be met—that is, the population standard deviations are not known. In these cases, a *t* test is used to test the difference between means when the two samples are independent and when the samples are taken from two normally or approximately normally distributed populations. Samples are **independent samples** when they are not related. Also it will be assumed that the variances are not equal.

Formula for the *t* Test—For Testing the Difference Between Two Means—Independent Samples

Variances are assumed to be unequal

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

where the degrees of freedom are equal to the smaller of $n_1 - 1$ or $n_2 - 1$.

The formula

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

follows the format of

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

where $\bar{X}_1 - \bar{X}_2$ is the observed difference between sample means and where the expected value $\mu_1 - \mu_2$ is equal to zero when no difference between population means is hypothesized. The denominator $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ is the standard error of the difference between two means. Since mathematical derivation of the standard error is somewhat complicated, it will be omitted here.

Confidence intervals can also be found for the difference between two means with this formula:

Confidence Intervals for the Difference of Two Means: Independent Samples

Variances assumed to be unequal:

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

d.f. = smaller value of $n_1 - 1$ or $n_2 - 1$

Example 9–5

Find the 95% confidence interval for the data in Example 9–4.

Solution

Substitute in the formula.

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(191 - 199) - 2.365\sqrt{\frac{38^2}{8} + \frac{12^2}{10}} < \mu_1 - \mu_2$$

$$< (191 - 199) + 2.365\sqrt{\frac{38^2}{8} + \frac{12^2}{10}}$$

$$-41.02 < \mu_1 - \mu_2 < 25.02$$

Since 0 is contained in the interval, the decision is to not reject the null hypothesis H_0: $\mu_1 = \mu_2$.

In many statistical software packages, a different method is used to compute the degrees of freedom for this *t* test. They are determined by the formula

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

This formula will not be used in this textbook.

There are actually two different options for the use of *t* tests. *One option is used when the variances of the populations are not equal, and the other option is used when the variances are equal.* To determine whether two sample variances are equal, the researcher can use an *F* test, as shown in Section 9–5.

When the variances are assumed to be equal, this formula is used and

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

follows the format of

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

For the numerator, the terms are the same as in the previously given formula. However, a note of explanation is needed for the denominator of the second test statistic. Since both