CHAPTER 10

Comparing Two Means and Two Proportions

Learning Objectives

After mastering the material in this chapter, you will be able to:

LO10-1 Compare two population means when the samples are independent.

LO10-2 Recognize when data come from independent samples and when they are paired.

LO10-3 Compare two population means when the data are paired.

LO10-4 Compare two population proportions using large independent samples.

Chapter Outline

Business improvement often requires making comparisons. For example, to increase consumer awareness of a product or service, it might be necessary to compare different types of advertising campaigns. Or to offer more profitable investments to its customers, an investment firm might compare the profitability of different investment portfolios. As a third example, a manufacturer might compare different production methods in order to minimize or eliminate out-of-specification product.

In this chapter we discuss using confidence intervals and hypothesis tests to **compare two populations.** Specifically, we compare two population means and two population proportions. We make these comparisons by studying **differences.** For instance, to compare two population means, say μ_1 and μ_2, we consider the difference between these means, $\mu_1 - \mu_2$. If, for example, we use a confidence interval or hypothesis test to conclude that $\mu_1 - \mu_2$ is a positive number, then we conclude that μ_1 is greater than μ_2. On the other hand, if a confidence interval or hypothesis test shows that $\mu_1 - \mu_2$ is a negative number, then we conclude that μ_1 is less than μ_2.

We explain many of this chapter's methods in the context of three new cases:

The Catalyst Comparison Case: The production supervisor at a chemical plant uses confidence intervals and hypothesis tests for the difference between two population means to determine which of two catalysts maximizes the hourly yield of a chemical process. By maximizing yield, the plant increases its productivity and improves its profitability.

The Auto Insurance Case: In order to reduce the costs of automobile accident claims, an insurance company uses confidence intervals and hypothesis tests for the difference between two population means to compare repair cost estimates for damaged cars at two different garages.

The Test Market Case: An advertising agency is test marketing a new product by using one advertising campaign in Des Moines, Iowa, and a different campaign in Toledo, Ohio. The agency uses confidence intervals and hypothesis tests for the difference between two population proportions to compare the effectiveness of the two advertising campaigns.

10.1 Comparing Two Population Means by Using Independent Samples

LO10-1 Compare two population means when the samples are independent.

A bank manager has developed a new system to reduce the time customers spend waiting to be served by tellers during peak business hours. We let μ_1 denote the population mean customer waiting time during peak business hours under the current system. To estimate μ_1, the manager randomly selects $n_1 = 100$ customers and records the length of time each customer spends waiting for service. The manager finds that the mean and the variance of the waiting times for these 100 customers are $\bar{x}_1 = 8.79$ minutes and $s_1^2 = 4.8237$. We let μ_2 denote the population mean customer waiting time during peak business hours for the new system. During a trial run, the manager finds that the mean and the variance of the waiting times for a random sample of $n_2 = 100$ customers are $\bar{x}_2 = 5.14$ minutes and $s_2^2 = 1.7927$.

In order to compare μ_1 and μ_2, the manager estimates $\mu_1 - \mu_2$, the difference between μ_1 and μ_2. Intuitively, a logical point estimate of $\mu_1 - \mu_2$ is the difference between the sample means

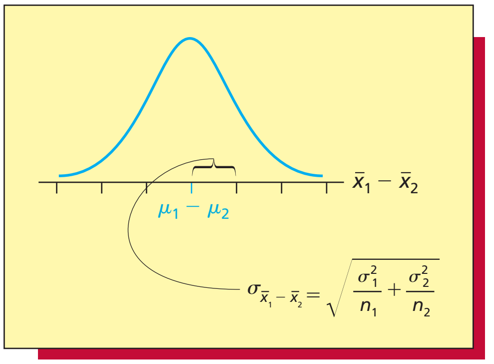
$$\bar{x}_1 - \bar{x}_2 = 8.79 - 5.14 = 3.65 \text{ minutes}$$

This says we estimate that the current population mean waiting time is 3.65 minutes longer than the population mean waiting time under the new system. That is, we estimate that the new system reduces the mean waiting time by 3.65 minutes.

To compute a confidence interval for $\mu_1 - \mu_2$ (or to test a hypothesis about $\mu_1 - \mu_2$), we need to know the properties of the sampling distribution of $\bar{x}_1 - \bar{x}_2$. To understand this sampling distribution, consider randomly selecting a sample[1] of n_1 measurements from a population having mean μ_1 and variance σ_1^2. Let $\bar{x}_1$ be the mean of this sample. Also consider randomly selecting a

[1]Each sample in this chapter is a *random* sample. As has been our practice throughout this book, for brevity we sometimes refer to "random samples" as "samples."

FIGURE 10.1 The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$ Has Mean $\mu_1 - \mu_2$ and Standard Deviation $\sigma_{\bar{x}_1 - \bar{x}_2}$



sample of n_2 measurements from another population having mean μ_2 and variance σ_2^2. Let $\bar{x}_2$ be the mean of this sample. Different samples from the first population would give different values of $\bar{x}_1$, and different samples from the second population would give different values of $\bar{x}_2$—so different pairs of samples from the two populations would give different values of $\bar{x}_1 - \bar{x}_2$. In the following box we describe the **sampling distribution of $\bar{x}_1 - \bar{x}_2$,** which is the probability distribution of all possible values of $\bar{x}_1 - \bar{x}_2$. Here we assume that the randomly selected samples from the two populations are independent of each other. This means that there is no relationship between the measurements in one sample and the measurements in the other sample. In such a case, we say that we are performing an **independent samples experiment.**

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

If the randomly selected samples are **independent** of each other, then the population of all possible values of $\bar{x}_1 - \bar{x}_2$

1. Has a normal distribution if each sampled population has a normal distribution, or has approximately a normal distribution if the sampled populations are not normally distributed and each of the sample sizes n_1 and n_2 is large.
2. Has mean $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$
3. Has standard deviation $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Figure 10.1 illustrates the sampling distribution of $\bar{x}_1 - \bar{x}_2$. Using this sampling distribution, we can find a confidence interval for and test a hypothesis about $\mu_1 - \mu_2$ by using the normal distribution. However, the interval and test assume that the true values of the population variances σ_1^2 and σ_2^2 are known, which is very unlikely. Therefore, we will estimate σ_1^2 and σ_2^2 by using s_1^2 and s_2^2, the variances of the samples randomly selected from the populations being compared, and base a confidence interval and a hypothesis test on the t distribution. There are two approaches to doing this. The first approach gives theoretically correct confidence intervals and hypothesis tests but assumes that the population variances σ_1^2 and σ_2^2 are equal. The second approach does not require that σ_1^2 and σ_2^2 are equal but gives only approximately correct confidence intervals and hypothesis tests. In the bank customer waiting time situation, the sample variances are $s_1^2 = 4.8237$ and $s_2^2 = 1.7927$. The difference in these sample variances makes it questionable to assume that the population variances are equal. More will be said later about deciding whether we can assume that two population variances are equal and about choosing between the two t-distribution approaches in a particular situation. For now, we will first consider the case where the population variances σ_1^2 and σ_2^2 can be assumed to be equal. Denoting the common value of these variances as σ^2, it follows that

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Because we are assuming that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we do not need separate estimates of σ_1^2 and σ_2^2. Instead, we combine the results of the two independent random samples to compute a single estimate of σ^2. This estimate is called the **pooled estimate** of σ^2, and it is a weighted average of the two sample variances s_1^2 and s_2^2. Denoting the pooled estimate as s_p^2, it is computed using the formula

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

Using s_p^2, the estimate of $\sigma_{\bar{x}_1-\bar{x}_2}$ is

$$\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and we form the statistic

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

It can be shown that, if we have randomly selected independent samples from two normally distributed populations having equal variances, then the sampling distribution of this statistic is a t distribution having $(n_1 + n_2 - 2)$ degrees of freedom. Therefore, we can obtain the following confidence interval for $\mu_1 - \mu_2$:

A *t*-Based Confidence Interval for the Difference between Two Population Means: Equal Variances

Suppose we have randomly selected independent samples from two normally distributed populations having equal variances. Then, a **100(1 − α) percent confidence interval for $\mu_1 - \mu_2$** is

$$\left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right] \quad \text{where} \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

and $t_{\alpha/2}$ is based on $(n_1 + n_2 - 2)$ degrees of freedom.

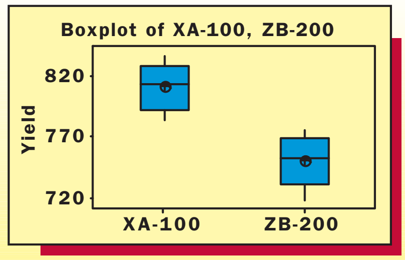
EXAMPLE 10.1 The Catalyst Comparison Case: Process Improvement

C

A production supervisor at a major chemical company must determine which of two catalysts, catalyst XA-100 or catalyst ZB-200, maximizes the hourly yield of a chemical process. In order to compare the mean hourly yields obtained by using the two catalysts, the supervisor runs the process using each catalyst for five one-hour periods. The resulting yields (in pounds per hour)

TABLE 10.1 Yields of a Chemical Process Obtained Using Two Catalysts DS Catalyst

Catalyst XA-100	Catalyst ZB-200
801	752
814	718
784	776
836	742
820	763
$\bar{x}_1 = 811$	$\bar{x}_2 = 750.2$
$s_1^2 = 386$	$s_2^2 = 484.2$



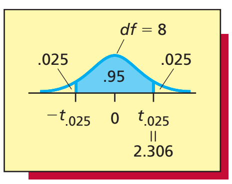
for each catalyst, along with the means, variances, and box plots[2] of the yields, are given in Table 10.1. Assuming that all other factors affecting yields of the process have been held as constant as possible during the test runs, it seems reasonable to regard the five observed yields for each catalyst as a random sample from the population of all possible hourly yields for the catalyst. Furthermore, because the sample variances $s_1^2 = 386$ and $s_2^2 = 484.2$ do not differ substantially (notice that $s_1 = 19.65$ and $s_2 = 22.00$ differ by even less), it might be reasonable to conclude that the population variances are approximately equal.[3] It follows that the pooled estimate

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(5 - 1)(386) + (5 - 1)(484.2)}{5 + 5 - 2} = 435.1$$

is a point estimate of the common variance σ^2.

We define μ_1 as the mean hourly yield obtained by using catalyst XA-100, and we define μ_2 as the mean hourly yield obtained by using catalyst ZB-200. If the populations of all possible hourly yields for the catalysts are normally distributed, then a 95 percent confidence interval for $\mu_1 - \mu_2$ is

$$\left[(\bar{x}_1 - \bar{x}_2) \pm t_{.025}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$
$$= \left[(811 - 750.2) \pm 2.306\sqrt{435.1\left(\frac{1}{5} + \frac{1}{5}\right)}\right]$$
$$= [60.8 \pm 30.4217]$$
$$= [30.38,\ 91.22]$$



Here $t_{.025} = 2.306$ is based on $n_1 + n_2 - 2 = 5 + 5 - 2 = 8$ degrees of freedom. This interval tells us that we are 95 percent confident that the mean hourly yield obtained by using catalyst XA-100 is between 30.38 and 91.22 pounds higher than the mean hourly yield obtained by using catalyst ZB-200.

Suppose we wish to test a hypothesis about $\mu_1 - \mu_2$. In the following box we describe how this can be done. Here we test the null hypothesis H_0: $\mu_1 - \mu_2 = D_0$, where D_0 is a number whose value varies depending on the situation. Often D_0 will be the number 0. In such a case, the null hypothesis H_0: $\mu_1 - \mu_2 = 0$ says there is **no difference** between the population means μ_1 and μ_2. In this case, each alternative hypothesis in the box implies that the population means μ_1 and μ_2 differ in a particular way.

[2]All of the box plots presented in this chapter and in Chapter 12 have been obtained using MINITAB.

[3]We describe how to test the equality of two variances in Chapter 11 (although, as we will explain, this test has drawbacks).

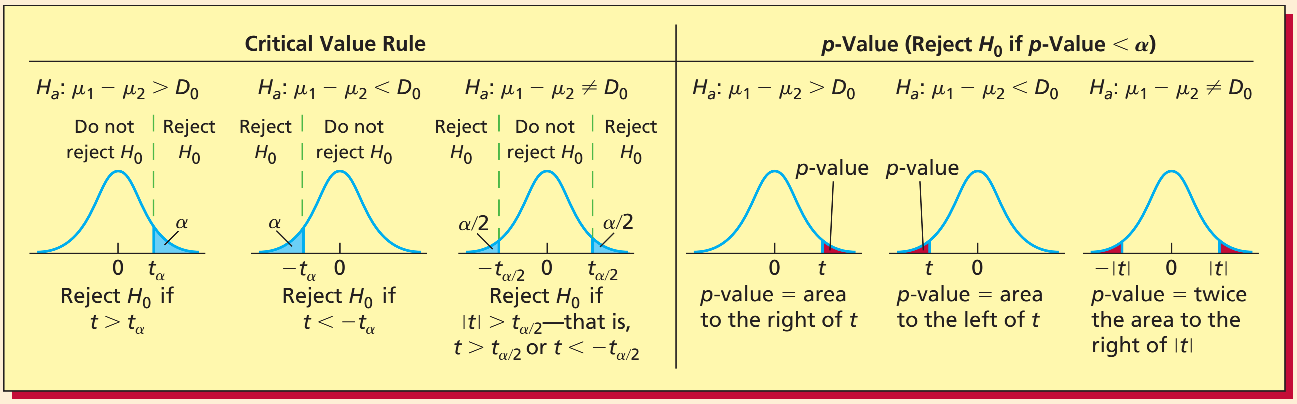
A t-Test about the Difference between Two Population Means: Equal Variances

Null Hypothesis $H_0: \mu_1 - \mu_2 = D_0$

Test Statistic $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Assumptions

Independent samples
and
Equal variances
and either
Normal populations
or
Large sample sizes

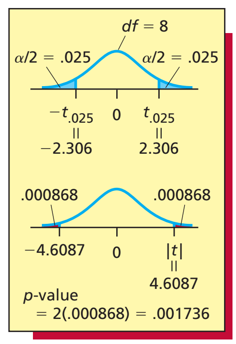


Here t_α, $t_{\alpha/2}$, and the p-values are based on $n_1 + n_2 - 2$ degrees of freedom.

EXAMPLE 10.2 The Catalyst Comparison Case: Process Improvement

C

In order to compare the mean hourly yields obtained by using catalysts XA-100 and ZB-200, we will test $\boldsymbol{H_0: \mu_1 - \mu_2 = 0}$ **versus** $\boldsymbol{H_a: \mu_1 - \mu_2 \neq 0}$ at the **.05 level of significance.** To perform the hypothesis test, we will use the sample information in Table 10.1 to calculate the value of the **test statistic t in the summary box.** Then, because $H_a: \mu_1 - \mu_2 \neq 0$ implies a two tailed test, we will **reject $H_0: \mu_1 - \mu_2 = 0$ if the absolute value of t is greater than $t_{\alpha/2} = t_{.025} = 2.306$.** Here the $t_{\alpha/2}$ point is based on $n_1 + n_2 - 2 = 5 + 5 - 2 = 8$ degrees of freedom. Using the data in Table 10.1, the **value of the test statistic** is



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(811 - 750.2) - 0}{\sqrt{435.1\left(\frac{1}{5} + \frac{1}{5}\right)}} = 4.6087$$

Because $|t|$ **= 4.6087 is greater than $t_{.025} = 2.306$, we can reject $H_0: \mu_1 - \mu_2 = 0$ in favor of $H_a: \mu_1 - \mu_2 \neq 0$.** We conclude (at an α of .05) that the mean hourly yields obtained by using the two catalysts differ. Furthermore, the point estimate $\bar{x}_1 - \bar{x}_2 = 811 - 750.2 = 60.8$ says we estimate that the mean hourly yield obtained by using catalyst XA-100 is 60.8 pounds higher than the mean hourly yield obtained by using catalyst ZB-200.

BI

Figure 10.2(a) gives the Excel output for using the equal variance t statistic to test H_0 versus H_a. The output tells us that $t = 4.6087$ and that the associated p-value is .001736. This very small p-value tells us that we have very strong evidence against $H_0: \mu_1 - \mu_2 = 0$ and in favor of $H_a: \mu_1 - \mu_2 \neq 0$. In other words, we have very strong evidence that the mean hourly yields obtained by using the two catalysts differ. (Note that in Figure 10.2(b) we give the Excel output for using an *unequal variances t statistic,* which is discussed on the following pages, to perform the hypothesis test.)

FIGURE 10.2 Excel Outputs for Testing the Equality of Means in the Catalyst Comparison Case

(a) The Excel Output Assuming Equal Variances

t-Test: Two-Sample Assuming Equal Variances

	XA-100	ZB-200
Mean	811	750.2
Variance	386	484.2
Observations	5	5
Pooled Variance	435.1	
Hypothesized Mean Diff	0	
df	8	
t Stat	4.608706	
P(T<=t) one-tail	0.000868	
t Critical one-tail	1.859548	
P(T<=t) two-tail	0.001736	
t Critical two-tail	2.306004	

(b) The Excel Output Assuming Unequal Variances

t-Test: Two-Sample Assuming Unequal Variances

	XA-100	ZB-200
Mean	811	750.2
Variance	386	484.2
Observations	5	5
Hypothesized Mean Diff	0	
df	8	
t Stat	4.608706	
P(T<=t) one-tail	0.000868	
t Critical one-tail	1.859548	
P(T<=t) two-tail	0.001736	
t Critical two-tail	2.306004	

When the sampled populations are normally distributed and the population variances σ_1^2 and σ_2^2 differ, the following can be shown.

t-Based Confidence Intervals for $\mu_1 - \mu_2$, and *t*-Tests about $\mu_1 - \mu_2$: Unequal Variances

1 When the sample sizes n_1 and n_2 are equal, the "equal variances" *t*-based confidence interval and hypothesis test given in the preceding two boxes are approximately valid even if the population variances σ_1^2 and σ_2^2 differ substantially. As a rough rule of thumb, if the larger sample variance is not more than three times the smaller sample variance when the sample sizes are equal, we can use the equal variances interval and test.

2 Suppose that the larger sample variance is more than three times the smaller sample variance when the sample sizes are equal or suppose that both the sample sizes and the sample variances differ substantially. Then, we can use an approximate procedure that is sometimes called an "unequal variances" procedure. This procedure says that an **approximate 100(1 − α) percent confidence interval for $\mu_1 - \mu_2$** is

$$\left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$$

Furthermore, we can test $H_0\colon \mu_1 - \mu_2 = D_0$ by using the test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

and by using the previously given critical value and *p*-value conditions.

For both the interval and the test, the degrees of freedom are equal to

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

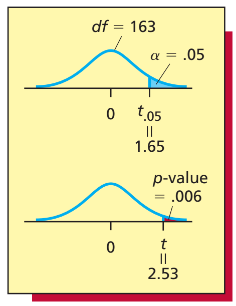
Here, if *df* is not a whole number, we can round *df* down to the next smallest whole number.

In general, both the "equal variances" and the "unequal variances" procedures have been shown to be approximately valid when the sampled populations are only approximately normally distributed (say, if they are mound-shaped). Furthermore, although the above summary box might seem to imply that we should use the unequal variances procedure only if we cannot use the equal variances procedure, this is not necessarily true. In fact, because the unequal variances procedure can be shown to be a very accurate approximation whether or not the population variances are equal and for most sample sizes (here, both n_1 and n_2 should be at least 5), **many statisticians believe that it is best to use the unequal variances procedure in almost every situation.** If each of n_1 and n_2 is large (at least 30), both the equal variances procedure and the unequal variances procedure are approximately valid, no matter what probability distributions describe the sampled populations.

To illustrate the unequal variances procedure, consider the bank customer waiting time situation, and recall that $\mu_1 - \mu_2$ is the difference between the mean customer waiting time under the current system and the mean customer waiting time under the new system. Because of cost considerations, the bank manager wants to implement the new system only if it reduces the mean waiting time by more than three minutes. Therefore, the manager will test the **null hypothesis H_0: $\mu_1 - \mu_2 = 3$ versus the alternative hypothesis H_a: $\mu_1 - \mu_2 > 3$.** If H_0 can be rejected in favor of H_a at the **.05 level of significance,** the manager will implement the new system. Recall that a random sample of $n_1 = 100$ waiting times observed under the current system gives a sample mean $\bar{x}_1 = 8.79$ and a sample variance $s_1^2 = 4.8237$. Also, recall that a random sample of $n_2 = 100$ waiting times observed during the trial run of the new system yields a sample mean $\bar{x}_2 = 5.14$ and a sample variance $s_2^2 = 1.7927$. Because each sample is large, we can use the **unequal variances test statistic t in the summary box.** The degrees of freedom for this statistic are

$$
\begin{aligned}
df &= \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} \\
&= \frac{[(4.8237/100) + (1.7927/100)]^2}{\dfrac{(4.8237/100)^2}{99} + \dfrac{(1.7927/100)^2}{99}} \\
&= 163.657
\end{aligned}
$$

which we will round down to 163. Therefore, because H_a: $\mu_1 - \mu_2 > 3$ implies a right tailed test, we will **reject H_0: $\mu_1 - \mu_2 = 3$ if the value of the test statistic t is greater than $t_\alpha = t_{.05} = 1.65$** (which is based on 163 degrees of freedom and has been found using a computer). Using the sample data, the **value of the test statistic** is

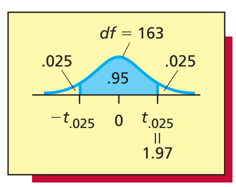


$$
t = \frac{(\bar{x}_1 - \bar{x}_2) - 3}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(8.79 - 5.14) - 3}{\sqrt{\dfrac{4.8237}{100} + \dfrac{1.7927}{100}}} = \frac{.65}{.25722} = 2.53
$$

Because $t = 2.53$ is greater than $t_{.05} = 1.65$, we reject H_0: $\mu_1 - \mu_2 = 3$ in favor of H_a: $\mu_1 - \mu_2 > 3$. We conclude (at an α of .05) that $\mu_1 - \mu_2$ is greater than 3 and, therefore, that the new system reduces the population mean customer waiting time by more than 3 minutes. Therefore, the bank manager will implement the new system. Furthermore, the point estimate $\bar{x}_1 - \bar{x}_2 = 3.65$ says that we estimate that the new system reduces mean waiting time by 3.65 minutes.

BI

Figure 10.3 gives the MINITAB output of using the unequal variances procedure to test H_0: $\mu_1 - \mu_2 = 3$ versus H_a: $\mu_1 - \mu_2 > 3$. The output tells us that $t = 2.53$ and that the associated p-value is .006. The very small p-value tells us that we have very strong evidence against H_0: $\mu_1 - \mu_2 = 3$ and in favor of H_a: $\mu_1 - \mu_2 > 3$. That is, we have very strong evidence that $\mu_1 - \mu_2$ is greater than 3 and, therefore, that the new system reduces the mean customer waiting time by more than 3 minutes. To find a 95 percent confidence interval for $\mu_1 - \mu_2$, note that we can use a computer to find that $t_{.025}$ based on 163 degrees of freedom is 1.97. It follows that the 95 percent confidence interval for $\mu_1 - \mu_2$ is



$$
\begin{aligned}
\left[(\bar{x}_1 - \bar{x}_2) \pm t_{.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] &= \left[(8.79 - 5.14) \pm 1.97\sqrt{\frac{4.8237}{100} + \frac{1.7927}{100}}\right] \\
&= [3.65 \pm .50792] \\
&= [3.14, 4.16]
\end{aligned}
$$

This interval says that we are 95 percent confident that the new system reduces the mean of all customer waiting times by between 3.14 minutes and 4.16 minutes.

FIGURE 10.3 MINITAB Output of the Unequal Variances Procedure for the Bank Customer Waiting Time Situation

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
Current	100	8.79	2.20	0.22
New	100	5.14	1.34	0.13

```
Difference = mu(1) - mu(2)
Estimate for difference:  3.650
95% lower bound for difference:  3.224
T-Test of difference = 3 (vs >):
T-Value = 2.53    P-Value = 0.006   DF = 163
```

FIGURE 10.4 MINITAB Output of the Unequal Variances Procedure for the Catalyst Comparison Case

Two-Sample T-Test and CI: XA-100, ZB-200

	N	Mean	StDev	SE Mean
XA-100	5	811.0	19.6	8.8
ZB-200	5	750.2	22.0	9.8

```
Difference = mu (XA-100) - mu (ZB-200)
Estimate for difference:  60.8000
95% CI for difference:  (29.6049, 91.9951)
T-Test of difference = 0 (vs not =):
  T-Value = 4.61    P-Value = 0.002   DF = 7
```

In general, the degrees of freedom for the unequal variances procedure will always be less than or equal to $n_1 + n_2 - 2$, the degrees of freedom for the equal variances procedure. For example, if we use the unequal variances procedure to analyze the catalyst comparison data in Table 10.1, we can calculate df to be 7.9. This is slightly less than $n_1 + n_2 - 2 = 5 + 5 - 2 = 8$, the degrees of freedom for the equal variances procedure. Figure 10.2(b) gives the Excel output, and Figure 10.4 gives the MINITAB output, of the unequal variances analysis of the catalyst comparison data. Note that the Excel unequal variances procedure rounds $df = 7.9$ up to 8 and obtains the same results as did the equal variances procedure (see Figure 10.2(a)). On the other hand, MINITAB rounds $df = 7.9$ down to 7 and finds that a 95 percent confidence interval for $\mu_1 - \mu_2$ is [29.6049, 91.9951]. MINITAB also finds that the test statistic for testing H_0: $\mu_1 - \mu_2 = 0$ versus H_a: $\mu_1 - \mu_2 \neq 0$ is $t = 4.61$ and that the associated p-value is .002. These results do not differ by much from the results given by the equal variances procedure.

To conclude this section, it is important to point out that if the sample sizes n_1 and n_2 are not large (at least 30), and if we fear that the sampled populations might be far from normally distributed, we can use a **nonparametric method.** One nonparametric method for comparing populations when using independent samples is the **Wilcoxon rank sum test.** This test is discussed in Chapter 18.

Exercises for Section 10.1

CONCEPTS

connect

For each of the formulas described below, list all of the assumptions that must be satisfied in order to validly use the formula.

10.1 The confidence interval in the formula box on page 383.

10.2 The hypothesis test described in the formula box on page 385.

10.3 The confidence interval and hypothesis test described in the formula box on page 386.

METHODS AND APPLICATIONS

Suppose we have taken independent, random samples of sizes $n_1 = 7$ and $n_2 = 7$ from two normally distributed populations having means μ_1 and μ_2, and suppose we obtain $\bar{x}_1 = 240$, $\bar{x}_2 = 210$, $s_1 = 5$, and $s_2 = 6$. Using the equal variances procedure, do Exercises 10.4, 10.5, and 10.6.

10.4 Calculate a 95 percent confidence interval for $\mu_1 - \mu_2$. Can we be 95 percent confident that $\mu_1 - \mu_2$ is greater than 20? Explain why we can use the equal variances procedure here.

10.5 Use critical values to test the null hypothesis H_0: $\mu_1 - \mu_2 \leq 20$ versus the alternative hypothesis H_a: $\mu_1 - \mu_2 > 20$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that the difference between μ_1 and μ_2 exceeds 20?

10.6 Use critical values to test the null hypothesis H_0: $\mu_1 - \mu_2 = 20$ versus the alternative hypothesis H_a: $\mu_1 - \mu_2 \neq 20$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that the difference between μ_1 and μ_2 is not equal to 20?

10.7 Repeat Exercises 10.4 through 10.6 using the unequal variances procedure. Compare your results to those obtained using the equal variances procedure.

10.8 An article in *Fortune* magazine reported on the rapid rise of fees and expenses charged by mutual funds. Assuming that stock fund expenses and municipal bond fund expenses are each approximately normally distributed, suppose a random sample of 12 stock funds gives a mean annual expense of 1.63 percent with a standard deviation of .31 percent, and an independent random sample of 12 municipal bond funds gives a mean annual expense of 0.89 percent with a standard deviation of .23 percent. Let μ_1 be the mean annual expense for stock funds, and let μ_2 be the mean annual expense for municipal bond funds. Do parts *a*, *b*, and *c* by using the equal variances procedure. Then repeat *a*, *b*, and *c* using the unequal variances procedure. Compare your results.

a Set up the null and alternative hypotheses needed to attempt to establish that the mean annual expense for stock funds is larger than the mean annual expense for municipal bond funds. Test these hypotheses at the .05 level of significance. What do you conclude?

b Set up the null and alternative hypotheses needed to attempt to establish that the mean annual expense for stock funds exceeds the mean annual expense for municipal bond funds by more than .5 percent. Test these hypotheses at the .05 level of significance. What do you conclude?

c Calculate a 95 percent confidence interval for the difference between the mean annual expenses for stock funds and municipal bond funds. Can we be 95 percent confident that the mean annual expense for stock funds exceeds that for municipal bond funds by more than .5 percent? Explain.

10.9 In the book *Business Research Methods,* Donald R. Cooper and C. William Emory (1995) discuss a manager who wishes to compare the effectiveness of two methods for training new salespeople. The authors describe the situation as follows:

> The company selects 22 sales trainees who are randomly divided into two experimental groups—one receives type *A* and the other type *B* training. The salespeople are then assigned and managed without regard to the training they have received. At the year's end, the manager reviews the performances of salespeople in these groups and finds the following results:

	***A* Group**	***B* Group**
Average Weekly Sales	$\bar{x}_1 = \$1,500$	$\bar{x}_2 = \$1,300$
Standard Deviation	$s_1 = 225$	$s_2 = 251$

a Set up the null and alternative hypotheses needed to attempt to establish that type *A* training results in higher mean weekly sales than does type *B* training.

b Because different sales trainees are assigned to the two experimental groups, it is reasonable to believe that the two samples are independent. Assuming that the normality assumption holds, and using the equal variances procedure, test the hypotheses you set up in part *a* at levels of significance .10, .05, .01, and .001. How much evidence is there that type *A* training produces results that are superior to those of type *B*?

c Use the equal variances procedure to calculate a 95 percent confidence interval for the difference between the mean weekly sales obtained when type *A* training is used and the mean weekly sales obtained when type *B* training is used. Interpret this interval.

10.10 A marketing research firm wishes to compare the prices charged by two supermarket chains—Miller's and Albert's. The research firm, using a standardized one-week shopping plan (grocery list), makes identical purchases at 10 of each chain's stores. The stores for each chain are randomly selected, and all purchases are made during a single week.

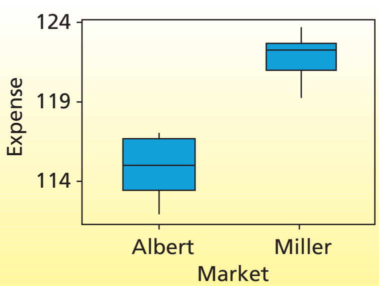
The shopping expenses obtained at the two chains, along with box plots of the expenses, are as follows: DS ShopExp

Miller's

\$119.25	\$121.32	\$122.34	\$120.14	\$122.19
\$123.71	\$121.72	\$122.42	\$123.63	\$122.44

Albert's

\$111.99	\$114.88	\$115.11	\$117.02	\$116.89
\$116.62	\$115.38	\$114.40	\$113.91	\$111.87



Because the stores in each sample are different stores in different chains, it is reasonable to assume that the samples are independent, and we assume that weekly expenses at each chain are normally distributed.

a Letting μ_M be the mean weekly expense for the shopping plan at Miller's, and letting μ_A be the mean weekly expense for the shopping plan at Albert's, Figure 10.5 gives the MINITAB output of the test of H_0: $\mu_M - \mu_A = 0$ (that is, there is no difference between μ_M and μ_A) versus H_a: $\mu_M - \mu_A \neq 0$ (that is, μ_M and μ_A differ). Note that MINITAB has employed the

FIGURE 10.5 MINITAB Output of Testing the Equality of Mean Weekly Expenses at Miller's and Albert's Supermarket Chains (for Exercise 10.10)

```
Two-sample T for Millers vs Alberts

            N     Mean    StDev   SE Mean
Millers    10   121.92     1.40      0.44
Alberts    10   114.81     1.84      0.58

Difference = mu(Millers)- mu(Alberts)   Estimate for difference: 7.10900
95% CI for difference:  (5.57350, 8.64450)
T-Test of diff = 0 (vs not =): T-Value = 9.73   P-Value = 0.000  DF = 18
Both use Pooled StDev = 1.6343
```

equal variances procedure. Use the sample data to show that $\bar{x}_M = 121.92$, $s_M = 1.40$, $\bar{x}_A = 114.81$, $s_A = 1.84$, and $t = 9.73$.

b Using the t statistic given on the output and critical values, test H_0 versus H_a by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean weekly expenses at Miller's and Albert's differ?

c Figure 10.5 gives the p-value for testing H_0: $\mu_M - \mu_A = 0$ versus H_a: $\mu_M - \mu_A \neq 0$. Use the p-value to test H_0 versus H_a by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean weekly expenses at Miller's and Albert's differ?

d Figure 10.5 gives a 95 percent confidence interval for $\mu_M - \mu_A$. Use this confidence interval to describe the size of the difference between the mean weekly expenses at Miller's and Albert's. Do you think that these means differ in a practically important way?

e Set up the null and alternative hypotheses needed to attempt to establish that the mean weekly expense for the shopping plan at Miller's exceeds the mean weekly expense at Albert's by more than $5. Test the hypotheses at the .10, .05, .01, and .001 levels of significance. How much evidence is there that the mean weekly expense at Miller's exceeds that at Albert's by more than $5?

10.11 A large discount chain compares the performance of its credit managers in Ohio and Illinois by comparing the mean dollar amounts owed by customers with delinquent charge accounts in these two states. Here a small mean dollar amount owed is desirable because it indicates that bad credit risks are not being extended large amounts of credit. Two independent, random samples of delinquent accounts are selected from the populations of delinquent accounts in Ohio and Illinois, respectively. The first sample, which consists of 10 randomly selected delinquent accounts in Ohio, gives a mean dollar amount of $524 with a standard deviation of $68. The second sample, which consists of 20 randomly selected delinquent accounts in Illinois, gives a mean dollar amount of $473 with a standard deviation of $22.

a Set up the null and alternative hypotheses needed to test whether there is a difference between the population mean dollar amounts owed by customers with delinquent charge accounts in Ohio and Illinois.

b Figure 10.6 gives the MINITAB output of using the unequal variances procedure to test the equality of mean dollar amounts owed by customers with delinquent charge accounts in Ohio and Illinois. Assuming that the normality assumption holds, test the hypotheses you set up in part *a* by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean dollar amounts owed in Ohio and Illinois differ?

c Assuming that the normality assumption holds, calculate a 95 percent confidence interval for the difference between the mean dollar amounts owed in Ohio and Illinois. Based on this interval, do you think that these mean dollar amounts differ in a practically important way?

10.12 A loan officer compares the interest rates for 48-month fixed-rate auto loans and 48-month variable-rate auto loans. Two independent, random samples of auto loan rates are selected. A sample of eight 48-month fixed-rate auto loans had the following loan rates: DS AutoLoan

4.29% 3.75% 3.50% 3.99% 3.75% 3.99% 5.40% 4.00%

while a sample of five 48-month variable-rate auto loans had loan rates as follows:

3.59% 2.75% 2.99% 2.50% 3.00%

a Set up the null and alternative hypotheses needed to determine whether the mean rates for 48-month fixed-rate and variable-rate auto loans differ.

b Figure 10.7 gives the Excel output of using the equal variances procedure to test the hypotheses you set up in part *a*. Assuming that the normality and equal variances assumptions hold, use the Excel output and critical values to test these hypotheses by setting α equal to

FIGURE 10.6 MINITAB Output of Testing the Equality of Mean Dollar Amounts Owed for Ohio and Illinois (for Exercise 10.11)

Two-Sample T-Test and CI

```
Sample     N    Mean    StDev   SE Mean
Ohio      10   524.0    68.0        22
Illinois  20   473.0    22.0       4.9

Difference = mu(1) - mu(2)
Estimate for difference:  51.0
95% CI for difference:  (1.1, 100.9)
T-Test of difference = 0 (vs not =):
  T-Value = 2.31    P-Value = 0.046   DF = 9
```

FIGURE 10.7 Excel Output of Testing the Equality of Mean Loan Rates for Fixed and Variable 48-Month Auto Loans (for Exercise 10.12)

t-Test: Two-Sample Assuming Equal Variances

	Fixed-Rate (%)	Variable-Rate (%)
Mean	4.0838	2.966
Variance	0.3376	0.1637
Observations	8	5
Pooled Variance	0.2744	
Hypothesized Mean Difference	0	
df	11	
t Stat	3.7431	
P(T<=t) one-tail	0.0016	
t Critical one-tail	1.7959	
P(T<=t) two-tail	0.0032	
t Critical two-tail	2.2010	

.10, .05, .01, and .001. How much evidence is there that the mean rates for 48-month fixed- and variable-rate auto loans differ?

c Figure 10.7 gives the *p*-value for testing the hypotheses you set up in part *a*. Use the *p*-value to test these hypotheses by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean rates for 48-month fixed- and variable-rate auto loans differ?

d Calculate a 95 percent confidence interval for the difference between the mean rates for fixed- and variable-rate 48-month auto loans. Can we be 95 percent confident that the difference between these means exceeds .4 percent? Explain.

e Use a hypothesis test to establish that the difference between the mean rates for fixed- and variable-rate 48-month auto loans exceeds .4 percent. Use α equal to .05.

10.2 Paired Difference Experiments

EXAMPLE 10.3 The Auto Insurance Case: Comparing Mean Repair Costs

C

LO10-2 Recognize when data come from independent samples and when they are paired.

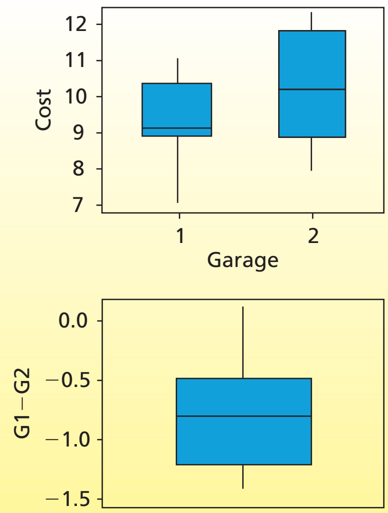
Home State Casualty, specializing in automobile insurance, wishes to compare the repair costs of moderately damaged cars (repair costs between \$700 and \$1,400) at two garages. One way to study these costs would be to take two independent samples (here we arbitrarily assume that each sample is of size $n = 7$). First we would randomly select seven moderately damaged cars that have recently been in accidents. Each of these cars would be taken to the first garage (garage 1), and repair cost estimates would be obtained. Then we would randomly select seven *different* moderately damaged cars, and repair cost estimates for these cars would be obtained at the second garage (garage 2). This sampling procedure would give us independent samples because the cars taken to garage 1 differ from those taken to garage 2. However, because the repair costs for moderately damaged cars can range from \$700 to \$1,400, there can be substantial differences in damages to moderately damaged cars. These differences might tend to conceal any real differences between repair costs at the two garages. For example, suppose the repair cost estimates for the cars taken to garage 1 are higher than those for the cars taken to garage 2. This difference might exist because garage 1 charges customers more for repair work than does garage 2. However, the difference could also arise because the cars taken to garage 1 are more severely damaged than the cars taken to garage 2.

To overcome this difficulty, we can perform a **paired difference experiment.** Here we could randomly select one sample of $n = 7$ moderately damaged cars. The cars in this sample would be taken to both garages, and a repair cost estimate for each car would be obtained at each garage. The advantage of the paired difference experiment is that the repair cost estimates at the two garages are obtained for the same cars. Thus, any true differences in the repair cost estimates would not be concealed by possible differences in the severity of damages to the cars.

Suppose that when we perform the paired difference experiment, we obtain the repair cost estimates in Table 10.2 (these estimates are given in units of \$100). To analyze these data, we

TABLE 10.2 **A Sample of n = 7 Paired Differences of the Repair Cost Estimates at Garages 1 and 2 (Cost Estimates in Hundreds of Dollars)** DS Repair

Sample of $n = 7$ Damaged Cars	Repair Cost Estimates at Garage 1	Repair Cost Estimates at Garage 2	Sample of $n = 7$ Paired Differences
Car 1	\$ 7.1	\$ 7.9	$d_1 = -.8$
Car 2	9.0	10.1	$d_2 = -1.1$
Car 3	11.0	12.2	$d_3 = -1.2$
Car 4	8.9	8.8	$d_4 = .1$
Car 5	9.9	10.4	$d_5 = -.5$
Car 6	9.1	9.8	$d_6 = -.7$
Car 7	10.3	11.7	$d_7 = -1.4$
	$\bar{x}_1 = 9.329$	$\bar{x}_2 = 10.129$	$\bar{d} = -.8 = \bar{x}_1 - \bar{x}_2$
			$s_d^2 = .2533$
			$s_d = .5033$



calculate the difference between the repair cost estimates at the two garages for each car. The resulting **paired differences** are given in the last column of Table 10.2. The mean of the sample of $n = 7$ paired differences is

$$\bar{d} = \frac{-.8 + (-1.1) + (-1.2) + \cdot \cdot \cdot + (-1.4)}{7} = -.8$$

which equals the difference between the sample means of the repair cost estimates at the two garages

$$\bar{x}_1 - \bar{x}_2 = 9.329 - 10.129 = -.8$$

Furthermore, $\bar{d} = -.8$ (that is, $-\$80$) is the point estimate of

$$\mu_d = \mu_1 - \mu_2$$

the mean of the population of all possible paired differences of the repair cost estimates (for all possible moderately damaged cars) at garages 1 and 2 (which is equivalent to μ_1, the mean of all possible repair cost estimates at garage 1, minus μ_2, the mean of all possible repair cost estimates at garage 2). This says we estimate that the mean of all possible repair cost estimates at garage 1 is \$80 less than the mean of all possible repair cost estimates at garage 2.

In addition, the variance and standard deviation of the sample of $n = 7$ paired differences

$$s_d^2 = \frac{\sum_{i=1}^{7} (d_i - \bar{d})^2}{7 - 1} = .2533$$

and

$$s_d = \sqrt{.2533} = .5033$$

are the point estimates of σ_d^2 and σ_d, the variance and standard deviation of the population of all possible paired differences.

LO10-3 Compare two population means when the data are paired.

In general, suppose we wish to compare two population means, μ_1 and μ_2. Also suppose that we have obtained two different measurements (for example, repair cost estimates) on the same n units (for example, cars), and suppose we have calculated the n paired differences between these measurements. Let $\bar{d}$ and s_d be the mean and the standard deviation of these n paired differences. If it is reasonable to assume that the paired differences have been randomly selected from a normally distributed (or at least mound-shaped) population of paired differences with mean μ_d and standard deviation σ_d, then the sampling distribution of

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

is a t distribution having $n - 1$ degrees of freedom. This implies that we have the following confidence interval for μ_d:

A Confidence Interval for the Mean, μ_d, of a Population of Paired Differences

Let μ_d be the mean of a **normally distributed population of paired differences,** and let $\bar{d}$ and s_d be the mean and standard deviation of a sample of n paired differences that have been randomly selected from the population. Then, a **100(1 − α) percent confidence interval for $\mu_d = \mu_1 - \mu_2$** is

$$\left[\bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}\right]$$

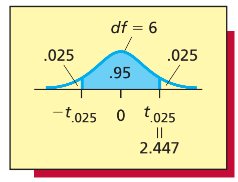
Here $t_{\alpha/2}$ is based on $(n - 1)$ degrees of freedom.

EXAMPLE 10.4 The Auto Insurance Case: Comparing Mean Repair Costs

C

Using the data in Table 10.2, and assuming that the population of paired repair cost differences is normally distributed, a 95 percent confidence interval for $\mu_d = \mu_1 - \mu_2$ is

$$\left[\bar{d} \pm t_{.025}\frac{s_d}{\sqrt{n}}\right] = \left[-.8 \pm 2.447\frac{.5033}{\sqrt{7}}\right]$$
$$= [-.8 \pm .4654]$$
$$= [-1.2654, -.3346]$$



Here $t_{.025} = 2.447$ is based on $n - 1 = 7 - 1 = 6$ degrees of freedom. This interval says that Home State Casualty can be 95 percent confident that μ_d, the mean of all possible paired differences of the repair cost estimates at garages 1 and 2, is between −\$126.54 and −\$33.46. That is, we are 95 percent confident that μ_1, the mean of all possible repair cost estimates at garage 1, is between \$126.54 and \$33.46 less than μ_2, the mean of all possible repair cost estimates at garage 2.

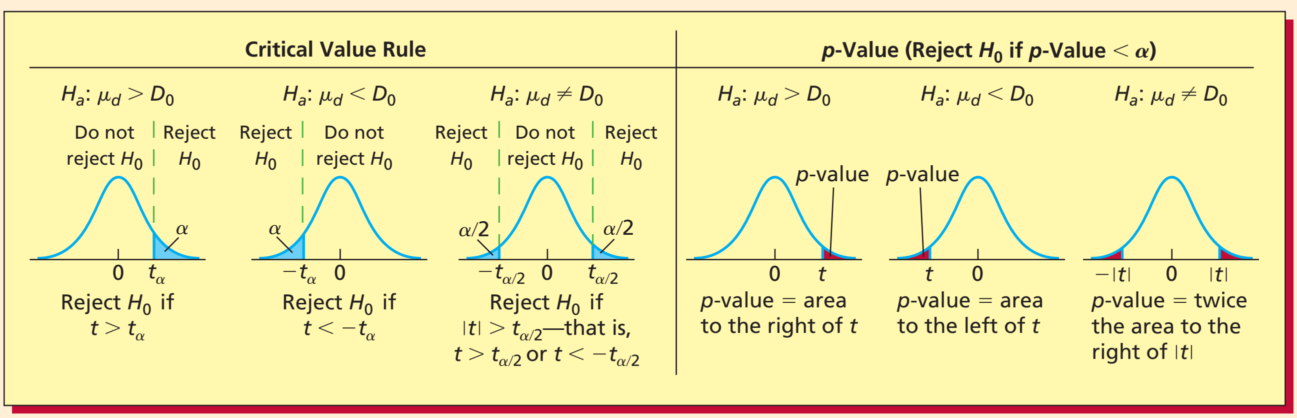
We can also test a hypothesis about μ_d, the mean of a population of paired differences. We show how to test the null hypothesis

$$H_0: \mu_d = D_0$$

in the following box. Here the value of the constant D_0 depends on the particular problem. Often D_0 equals 0, and the null hypothesis $H_0: \mu_d = 0$ says that μ_1 and μ_2 do not differ.

Testing a Hypothesis about the Mean, μ_d, of a Population of Paired Differences

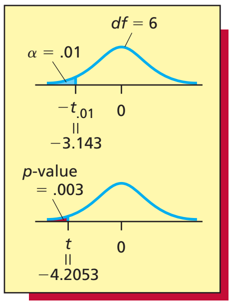
Null Hypothesis $H_0: \mu_d = D_0$ **Test Statistic** $t = \dfrac{\bar{d} - D_0}{s_d/\sqrt{n}}$ $df = n - 1$ **Assumptions** Normal population of paired differences or Large sample size



EXAMPLE 10.5 The Auto Insurance Case: Comparing Mean Repair Costs

C

Home State Casualty currently contracts to have moderately damaged cars repaired at garage 2. However, a local insurance agent suggests that garage 1 provides less expensive repair service that is of equal quality. Because it has done business with garage 2 for years, Home State has decided to give some of its repair business to garage 1 only if it has very strong evidence that μ_1, the mean repair cost estimate at garage 1, is smaller than μ_2, the mean repair cost estimate at garage 2—that is, if $\mu_d = \mu_1 - \mu_2$ is less than zero. Therefore, we will test **H_0: $\mu_d = 0$ or, equivalently, H_0: $\mu_1 - \mu_2 = 0$, versus H_a: $\mu_d < 0$ or, equivalently, H_a: $\mu_1 - \mu_2 < 0$,** at the **.01 level of significance.** To perform the hypothesis test, we will use the sample data in Table 10.2 to calculate the value of the **test statistic t in the summary box.** Because H_a: $\mu_d < 0$ implies a left tailed test, we will **reject H_0: $\mu_d = 0$ if the value of t is less than $-t_\alpha = -t_{.01} = -3.143$.** Here the t_α point is based on $n - 1 = 7 - 1 = 6$ degrees of freedom. Using the data in Table 10.2, the **value of the test statistic** is



$$t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{-.8 - 0}{.5033/\sqrt{7}} = -4.2053$$

Because $t = -4.2053$ is less than $-t_{.01} = -3.143$, we can reject H_0: $\mu_d = 0$ in favor of H_a: $\mu_d < 0$. We conclude (at an α of .01) that μ_1, the mean repair cost estimate at garage 1, is less than μ_2, the mean repair cost estimate at garage 2. As a result, Home State will give some of its repair business to garage 1. Furthermore, Figure 10.8 gives the MINITAB output of this hypothesis test and shows us that the p-value for the test is .003. Because this p-value is very small, we have very strong evidence that H_0 should be rejected and that μ_1 is less than μ_2.

BI

Figure 10.9 shows the Excel output for testing H_0: $\mu_d = 0$ versus H_a: $\mu_d < 0$ (the "one-tail" test) and for testing H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$ (the "two-tail" test). The Excel p-value for testing H_0: $\mu_d = 0$ versus H_a: $\mu_d < 0$ is .002826, which in the rounded form .003 is the same as

FIGURE 10.8 MINITAB Output of Testing H_0: $\mu_d = 0$ versus H_a: $\mu_d < 0$

```
Paired T for Garage1 - Garage2

              N        Mean      StDev    SE Mean
Garage1       7      9.3286     1.2500     0.4724
Garage2       7     10.1286     1.5097     0.5706
Difference    7   -0.800000   0.503322   0.190238

T-Test of mean difference = 0 (vs < 0):
                          T-Value = -4.21     P-Value = 0.003
```

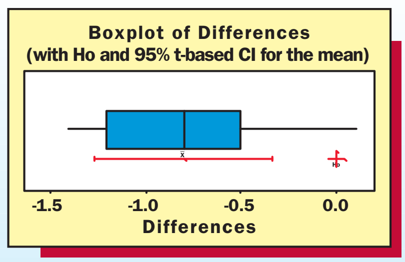


FIGURE 10.9 Excel Output of Testing H_0: $\mu_d = 0$

t-Test: Paired Two Sample for Means

	Garage1	Garage2
Mean	9.328571	10.12857
Variance	1.562381	2.279048
Observations	7	7
Pearson Correlation	0.950744	
Hypothesized Mean	0	
df	6	
t Stat	−4.20526	
P(T<=t) one-tail	0.002826	
t Critical one-tail	1.943181	
P(T<=t) two-tail	0.005653	
t Critical two-tail	2.446914	

the MINITAB p-value. This very small p-value tells us that Home State has very strong evidence that the mean repair cost at garage 1 is less than the mean repair cost at garage 2. The Excel p-value for testing H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$ is .005653.

In general, an experiment in which we have obtained two different measurements on the same n units is called a **paired difference experiment.** The idea of this type of experiment is to remove the variability due to the variable (for example, the amount of damage to a car) on which the observations are paired. In many situations, a paired difference experiment will provide more information than an independent samples experiment. As another example, suppose that we wish to assess which of two different machines produces a higher hourly output. If we randomly select 10 machine operators and randomly assign 5 of these operators to test machine 1 and the others to test machine 2, we would be performing an independent samples experiment. This is because different machine operators test machines 1 and 2. However, any difference in machine outputs could be obscured by differences in the abilities of the machine operators. For instance, if the observed hourly outputs are higher for machine 1 than for machine 2, we might not be able to tell whether this is due to (1) the superiority of machine 1 or (2) the possible higher skill level of the operators who tested machine 1. Because of this, it might be better to randomly select five machine operators, thoroughly train each operator to use both machines, and have each operator test both machines. We would then be **pairing on the machine operator,** and this would remove the variability due to the differing abilities of the operators.

The formulas we have given for analyzing a paired difference experiment are based on the t distribution. These formulas assume that the population of all possible paired differences is normally distributed (or at least mound-shaped). If the sample size is large (say, at least 30), the t-based interval and tests of this section are approximately valid no matter what the shape of the population of all possible paired differences. If the sample size is small, and if we fear that the population of all paired differences might be far from normally distributed, we can use a nonparametric method. One nonparametric method for comparing two populations when using a paired difference experiment is the **Wilcoxon signed ranks test.** This nonparametric test is discussed in Chapter 18.

Exercises for Section 10.2

CONCEPTS

10.13 Explain how a paired difference experiment differs from an independent samples experiment in terms of how the data for these experiments are collected.

10.14 Why is a paired difference experiment sometimes more informative than an independent samples experiment? Give an example of a situation in which a paired difference experiment might be advantageous.

10.15 Suppose a company wishes to compare the hourly output of its employees before and after vacations. Explain how you would collect data for a paired difference experiment to make this comparison.

METHODS AND APPLICATIONS

10.16 Suppose a sample of 49 paired differences that have been randomly selected from a normally distributed population of paired differences yields a sample mean of $\bar{d} = 5$ and a sample standard deviation of $s_d = 7$.

- **a** Calculate a 95 percent confidence interval for $\mu_d = \mu_1 - \mu_2$. Can we be 95 percent confident that the difference between μ_1 and μ_2 is not equal to 0?
- **b** Test the null hypothesis H_0: $\mu_d = 0$ versus the alternative hypothesis H_a: $\mu_d \neq 0$ by setting α equal to .10, .05, .01, and .001. How much evidence is there that μ_d differs from 0? What does this say about how μ_1 and μ_2 compare?
- **c** The p-value for testing H_0: $\mu_d \leq 3$ versus H_a: $\mu_d > 3$ equals .0256. Use the p-value to test these hypotheses with α equal to .10, .05, .01, and .001. How much evidence is there that μ_d exceeds 3? What does this say about the size of the difference between μ_1 and μ_2?

10.17 Suppose a sample of 11 paired differences that has been randomly selected from a normally distributed population of paired differences yields a sample mean of $\bar{d} = 103.5$ and a sample standard deviation of $s_d = 5$.

a Calculate 95 percent and 99 percent confidence intervals for $\mu_d = \mu_1 - \mu_2$.

b Test the null hypothesis H_0: $\mu_d \leq 100$ versus H_a: $\mu_d > 100$ by setting α equal to .05 and .01. How much evidence is there that $\mu_d = \mu_1 - \mu_2$ exceeds 100?

c Test the null hypothesis H_0: $\mu_d \geq 110$ versus H_a: $\mu_d < 110$ by setting α equal to .05 and .01. How much evidence is there that $\mu_d = \mu_1 - \mu_2$ is less than 110?

10.18 In the book *Essentials of Marketing Research,* William R. Dillon, Thomas J. Madden, and Neil H. Firtle (1993) present preexposure and postexposure attitude scores from an advertising study involving 10 respondents. The data for the experiment are given in Table 10.3. Assuming that the differences between pairs of postexposure and preexposure scores are normally distributed: DS AdStudy

a Set up the null and alternative hypotheses needed to attempt to establish that the advertisement increases the mean attitude score (that is, that the mean postexposure attitude score is higher than the mean preexposure attitude score).

b Test the hypotheses you set up in part *a* at the .10, .05, .01, and .001 levels of significance. How much evidence is there that the advertisement increases the mean attitude score?

c Estimate the minimum difference between the mean postexposure attitude score and the mean preexposure attitude score. Justify your answer.

10.19 National Paper Company must purchase a new machine for producing cardboard boxes. The company must choose between two machines. The machines produce boxes of equal quality, so the company will choose the machine that produces (on average) the most boxes. It is known that there are substantial differences in the abilities of the company's machine operators. Therefore National Paper has decided to compare the machines using a paired difference experiment. Suppose that eight randomly selected machine operators produce boxes for one hour using machine 1 and for one hour using machine 2, with the following results: DS BoxYield

	Machine Operator							
	1	2	3	4	5	6	7	8
Machine 1	53	60	58	48	46	54	62	49
Machine 2	50	55	56	44	45	50	57	47

a Assuming normality, perform a hypothesis test to determine whether there is a difference between the mean hourly outputs of the two machines. Use $\alpha = .05$.

b Estimate the minimum and maximum differences between the mean outputs of the two machines. Justify your answer.

10.20 During 2011 a company implemented a number of policies aimed at reducing the ages of its customers' accounts. In order to assess the effectiveness of these measures, the company randomly selects 10 customer accounts. The average age of each account is determined for the years 2010 and 2011. These data are given in Table 10.4. Assuming that the population of paired differences between the average ages in 2011 and 2010 is normally distributed: DS AcctAge

TABLE 10.3 Preexposure and Postexposure Attitude Scores (for Exercise 10.18) DS AdStudy

Subject	Preexposure Attitudes (A_1)	Postexposure Attitudes (A_2)	Attitude Change (d_i)
1	50	53	3
2	25	27	2
3	30	38	8
4	50	55	5
5	60	61	1
6	80	85	5
7	45	45	0
8	30	31	1
9	65	72	7
10	70	78	8

Source: W. R. Dillon, T. J. Madden, and N. H. Firtle, *Essentials of Marketing Research* (Burr Ridge, IL: Richard D. Irwin, 1993), p. 435. Copyright © 1993. Reprinted by permission of McGraw-Hill Companies, Inc.

TABLE 10.4 Average Account Ages in 2010 and 2011 for 10 Randomly Selected Accounts (for Exercise 10.20) DS AcctAge

Account	Average Age of Account in 2011 (Days)	Average Age of Account in 2010 (Days)
1	27	35
2	19	24
3	40	47
4	30	28
5	33	41
6	25	33
7	31	35
8	29	51
9	15	18
10	21	28

FIGURE 10.10 Excel Output of a Paired Difference Analysis of the Account Age Data (for Exercise 10.20)

t-Test: Paired Two Sample for Means

	2011 Age	2010 Age
Mean	27	34
Variance	53.55556	104.2222
Observations	10	10
Pearson Correlation	0.804586	
Hypothesized Mean	0	
df	9	
t Stat	−3.61211	
P(T<=t) one-tail	0.00282	
t Critical one-tail	1.833114	
P(T<=t) Two-tail	0.005641	
t Critical two-tail	2.262159	

TABLE 10.5 Weekly Study Time Data for Students Who Perform Well on the MidTerm DS StudyTime

Students	1	2	3	4	5	6	7	8
Before	15	14	17	17	19	14	13	16
After	9	9	11	10	19	10	14	10

a Set up the null and alternative hypotheses needed to establish that the mean average account age has been reduced by the company's new policies.

b Figure 10.10 gives the Excel output needed to test the hypotheses of part *a*. Use critical values to test these hypotheses by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean average account age has been reduced?

c Figure 10.10 gives the *p*-value for testing the hypotheses of part *a*. Use the *p*-value to test these hypotheses by setting α equal to .10, .05, .01, and .001. How much evidence is there that the mean average account age has been reduced?

d Calculate a 95 percent confidence interval for the mean difference in the average account ages between 2011 and 2010. Estimate the minimum reduction in the mean average account ages from 2010 to 2011.

10.21 Do students reduce study time in classes where they achieve a higher midterm score? In a *Journal of Economic Education* article (Winter 2005), Gregory Krohn and Catherine O'Connor studied student effort and performance in a class over a semester. In an intermediate macroeconomics course, they found that "students respond to higher midterm scores by reducing the number of hours they subsequently allocate to studying for the course."[4] Suppose that a random sample of $n = 8$ students who performed well on the midterm exam was taken and weekly study times before and after the exam were compared. The resulting data are given in Table 10.5. Assume that the population of all possible paired differences is normally distributed.

a Set up the null and alternative hypotheses to test whether there is a difference in the population mean study time before and after the midterm exam.

b Below we present the MINITAB output for the paired differences test. Use the output and critical values to test the hypotheses at the .10, .05, and .01 levels of significance. Has the population mean study time changed?

Paired T-Test and CI: StudyBefore, StudyAfter

```
Paired T for StudyBefore - StudyAfter
              N     Mean     StDev  SE Mean
StudyBefore   8  15.6250    1.9955   0.7055
StudyAfter    8  11.5000    3.4226   1.2101
Difference    8  4.12500   2.99702  1.05961

95% CI for mean difference: (1.61943, 6.63057)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.89  P-Value = 0.006
```

c Use the *p*-value to test the hypotheses at the .10, .05, and .01 levels of significance. How much evidence is there against the null hypothesis?

[4]Source: "Student Effort and Performance over the Semester," *Journal of Economic Education*, Winter 2005, pages 3–28.

10.3 Comparing Two Population Proportions by Using Large, Independent Samples

EXAMPLE 10.6 The Test Market Case: Comparing Advertising Media

C

LO10-4 Compare two population proportions using large independent samples.

Suppose a new product was test marketed in the Des Moines, Iowa, and Toledo, Ohio, metropolitan areas. Equal amounts of money were spent on advertising in the two areas. However, different advertising media were employed in the two areas. Advertising in the Des Moines area was done entirely on television, while advertising in the Toledo area consisted of a mixture of television, radio, newspaper, and magazine ads. Two months after the advertising campaigns commenced, surveys are taken to estimate consumer awareness of the product. In the Des Moines area, 631 out of 1,000 randomly selected consumers are aware of the product, while in the Toledo area 798 out of 1,000 randomly selected consumers are aware of the product. We define p_1 to be the proportion of all consumers in the Des Moines area who are aware of the product and p_2 to be the proportion of all consumers in the Toledo area who are aware of the product. It follows that, because the sample proportions of consumers who are aware of the product in the Des Moines and Toledo areas are

$$\hat{p}_1 = \frac{631}{1{,}000} = .631$$

and

$$\hat{p}_2 = \frac{798}{1{,}000} = .798$$

then a point estimate of $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 = .631 - .798 = -.167$$

This says we estimate that p_1 is .167 less than p_2. That is, we estimate that the percentage of all consumers who are aware of the product in the Toledo area is 16.7 percentage points higher than the percentage in the Des Moines area.

In order to find a confidence interval for and to carry out a hypothesis test about $p_1 - p_2$, we need to know the properties of the **sampling distribution of $\hat{p}_1 - \hat{p}_2$.** In general, therefore, consider randomly selecting n_1 elements from a population, and assume that a proportion p_1 of all the elements in the population fall into a particular category. Let $\hat{p}_1$ denote the proportion of elements in the sample that fall into the category. Also, consider randomly selecting a sample of n_2 elements from a second population, and assume that a proportion p_2 of all the elements in this population fall into the particular category. Let $\hat{p}_2$ denote the proportion of elements in the second sample that fall into the category.

The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

If the randomly selected samples are independent of each other, then the population of all possible values of $\hat{p}_1 - \hat{p}_2$:

1. Approximately has a normal distribution if each of the sample sizes n_1 and n_2 is large. Here n_1 and n_2 are large enough if n_1p_1, $n_1(1 - p_1)$, n_2p_2, and $n_2(1 - p_2)$ are all at least 5.
2. Has mean $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$
3. Has standard deviation $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{p_1(1 - p_1)}{n_1} + \dfrac{p_2(1 - p_2)}{n_2}}$

If we estimate p_1 by $\hat{p}_1$ and p_2 by $\hat{p}_2$ in the expression for $\sigma_{\hat{p}_1 - \hat{p}_2}$, then the sampling distribution of $\hat{p}_1 - \hat{p}_2$ implies the following $100(1 - \alpha)$ percent confidence interval for $p_1 - p_2$.

A Large Sample Confidence Interval for the Difference between Two Population Proportions[5]

Suppose we randomly select a sample of size n_1 from a population, and let $\hat{p}_1$ denote the proportion of elements in this sample that fall into a category of interest. Also suppose we randomly select a sample of size n_2 from another population, and let $\hat{p}_2$ denote the proportion of elements in this second sample that fall into the category of interest. Then, if each of the sample sizes n_1 and n_2 is large (n_1 and n_2 are considered to be large if $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, and $n_2(1-\hat{p}_2)$ are all at least 5), and if the random samples are independent of each other, a **100(1 − α) percent confidence interval for $p_1 - p_2$** is

$$\left[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

EXAMPLE 10.7 The Test Market Case: Comparing Advertising Media

C

Recall that in the advertising media situation described at the beginning of this section, 631 of 1,000 randomly selected consumers in Des Moines are aware of the new product, while 798 of 1,000 randomly selected consumers in Toledo are aware of the new product. Also recall that

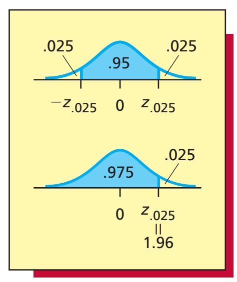
$$\hat{p}_1 = \frac{631}{1{,}000} = .631$$

and

$$\hat{p}_2 = \frac{798}{1{,}000} = .798$$

Because $n_1\hat{p}_1 = 1{,}000(.631) = 631$, $n_1(1-\hat{p}_1) = 1{,}000(1-.631) = 369$, $n_2\hat{p}_2 = 1{,}000(.798) = 798$, and $n_2(1-\hat{p}_2) = 1{,}000(1-.798) = 202$ are all at least 5, both n_1 and n_2 can be considered large. It follows that a 95 percent confidence interval for $p_1 - p_2$ is

$$\begin{aligned}
&\left[(\hat{p}_1 - \hat{p}_2) \pm z_{.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right] \\
&= \left[(.631 - .798) \pm 1.96\sqrt{\frac{(.631)(.369)}{1{,}000} + \frac{(.798)(.202)}{1{,}000}}\right] \\
&= [-.167 \pm .0389] \\
&= [-.2059, \ -.1281]
\end{aligned}$$



This interval says we are 95 percent confident that p_1, the proportion of all consumers in the Des Moines area who are aware of the product, is between .2059 and .1281 less than p_2, the proportion of all consumers in the Toledo area who are aware of the product. Thus, we have substantial evidence that advertising the new product by using a mixture of television, radio, newspaper, and magazine ads (as in Toledo) is more effective than spending an equal amount of money on television commercials only.

BI

[5] More correctly, because $\hat{p}_1(1-\hat{p}_1)/(n_1-1)$ and $\hat{p}_2(1-\hat{p}_2)/(n_2-1)$ are unbiased point estimates of $p_1(1-p_1)/n_1$ and $p_2(1-p_2)/n_2$, a point estimate of $\sigma_{\hat{p}_1-\hat{p}_2}$ is

$$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1}}$$

and a 100(1 − α) percent confidence interval for $p_1 - p_2$ is $[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}s_{\hat{p}_1-\hat{p}_2}]$. Because both n_1 and n_2 are large, there is little difference between the interval obtained by using this formula and those obtained by using the formula in the box above.

To test the null hypothesis H_0: $p_1 - p_2 = D_0$, we use the test statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$$

A commonly employed special case of this hypothesis test is obtained by setting D_0 equal to 0. In this case, the null hypothesis H_0: $p_1 - p_2 = 0$ says there is **no difference** between the population proportions p_1 and p_2. When $D_0 = 0$, the best estimate of the common population proportion $p = p_1 = p_2$ is obtained by computing

$$\hat{p} = \frac{\text{the total number of elements in the two samples that fall into the category of interest}}{\text{the total number of elements in the two samples}}$$

Therefore, the point estimate of $\sigma_{\hat{p}_1 - \hat{p}_2}$ is

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

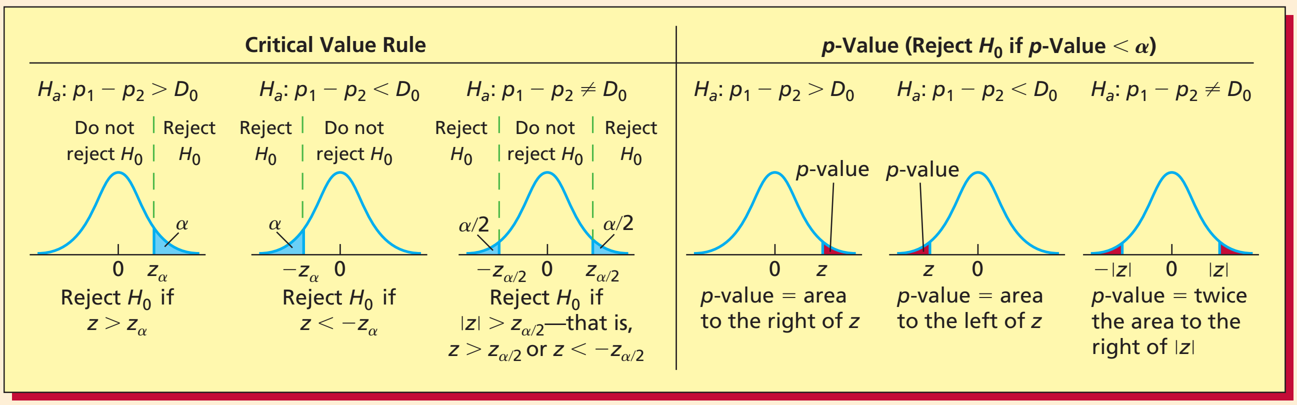
For the case where $D_0 \neq 0$, the point estimate of $\sigma_{\hat{p}_1 - \hat{p}_2}$ is obtained by estimating p_1 by $\hat{p}_1$ and p_2 by $\hat{p}_2$. With these facts in mind, we present the following procedure for testing H_0: $p_1 - p_2 = D_0$:

A Hypothesis Test about the Difference between Two Population Proportions

Null Hypothesis H_0: $p_1 - p_2 = D_0$

Test Statistic $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}}$

Assumptions Independent samples and Large sample sizes



Note:

1 If $D_0 = 0$, we estimate $\sigma_{\hat{p}_1 - \hat{p}_2}$ by

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

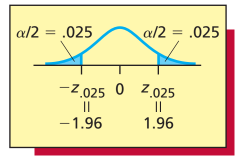
2 If $D_0 \neq 0$, we estimate $\sigma_{\hat{p}_1 - \hat{p}_2}$ by

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

EXAMPLE 10.8 The Test Market Case: Comparing Advertising Media

C

Recall that p_1 is the proportion of all consumers in the Des Moines area who are aware of the new product and that p_2 is the proportion of all consumers in the Toledo area who are aware of the new product. To test for the equality of these proportions, we will test $\boldsymbol{H_0: p_1 - p_2 = 0}$ **versus** $\boldsymbol{H_a: p_1 - p_2 \neq 0}$ at the **.05 level of significance.** Because both of the Des Moines and Toledo samples are large (see Example 10.7), we will calculate the value of the **test statistic z in the summary box** (where $D_0 = 0$). Since H_a: $p_1 - p_2 \neq 0$ implies a two tailed test, we will **reject $\boldsymbol{H_0: p_1 - p_2 = 0}$ if the absolute value of z is greater than $\boldsymbol{z_{\alpha/2} = z_{.05/2} = z_{.025} = 1.96}$.** Because 631 out of 1,000 randomly selected Des Moines residents were aware of the product and 798 out of 1,000 randomly selected Toledo residents were aware of the product, the estimate of $p = p_1 = p_2$ is



$$\hat{p} = \frac{631 + 798}{1{,}000 + 1{,}000} = \frac{1{,}429}{2{,}000} = .7145$$

and the **value of the test statistic is**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(.631 - .798) - 0}{\sqrt{(.7145)(.2855)\left(\frac{1}{1{,}000} + \frac{1}{1{,}000}\right)}} = \frac{-.167}{.0202} = -8.2673$$

Because $|z|$ = 8.2673 is greater than 1.96, we can reject $\boldsymbol{H_0: p_1 - p_2 = 0}$ in favor of $\boldsymbol{H_a: p_1 - p_2 \neq 0}$. We conclude (at an α of .05) that the proportions of all consumers who are aware of the product in Des Moines and Toledo differ. Furthermore, the point estimate $\hat{p}_1 - \hat{p}_2 = .631 - .798 = -.167$ says we estimate that the percentage of all consumers who are aware of the product in Toledo is 16.7 percentage points higher than the percentage of all consumers who are aware of the product in Des Moines. The p-value for this test is twice the area under the standard normal curve to the right of $|z| = 8.2673$. Because the area under the standard normal curve to the right of 3.99 is .00003, the p-value for testing H_0 is less than 2(.00003) = .00006. It follows that we have extremely strong evidence that H_0: $p_1 - p_2 = 0$ should be rejected in favor of H_a: $p_1 - p_2 \neq 0$. That is, this small p-value provides extremely strong evidence that p_1 and p_2 differ. Figure 10.11 presents the MINITAB output of the hypothesis test of H_0: $p_1 - p_2 = 0$ versus H_a: $p_1 - p_2 \neq 0$ and of a 95 percent confidence interval for $p_1 - p_2$. Note that the MINITAB output gives a value of the test statistic z (that is, the value −8.41) that is slightly different from the value −8.2673 calculated above. The reason is that, even though we are testing H_0: $p_1 - p_2 = 0$, MINITAB uses the second formula in the summary box (rather than the first formula) to calculate $s_{\hat{p}_1 - \hat{p}_2}$.

BI

FIGURE 10.11 MINITAB Output of Statistical Inference in the Test Market Case

Test and CI for Two Proportions

```
Sample     X      N   Sample p
1        631   1000   0.631000
2        798   1000   0.798000

Difference = p(1) - p(2)
Estimate for difference: -0.167
95% CI for difference: (-0.205906, -0.128094)
Test of difference = 0 (vs not = 0): Z = -8.41, P-value = 0.000
```

Exercises for Section 10.3

CONCEPTS

10.22 Explain what population is described by the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

10.23 What assumptions must be satisfied in order to use the methods presented in this section?

METHODS AND APPLICATIONS

In Exercises 10.24 through 10.26 we assume that we have selected two independent random samples from populations having proportions p_1 and p_2 and that $\hat{p}_1 = 800/1{,}000 = .8$ and $\hat{p}_2 = 950/1{,}000 = .95$.

10.24 Calculate a 95 percent confidence interval for $p_1 - p_2$. Interpret this interval. Can we be 95 percent confident that $p_1 - p_2$ is less than 0? That is, can we be 95 percent confident that p_1 is less than p_2? Explain.

10.25 Test H_0: $p_1 - p_2 = 0$ versus H_a: $p_1 - p_2 \neq 0$ by using critical values and by setting α equal to .10, .05, .01, and .001. How much evidence is there that p_1 and p_2 differ? Explain.

10.26 Test H_0: $p_1 - p_2 \geq -.12$ versus H_a: $p_1 - p_2 < -.12$ by using a p-value and by setting α equal to .10, .05, .01, and .001. How much evidence is there that p_2 exceeds p_1 by more than .12? Explain.

10.27 In an article in the *Journal of Advertising,* Weinberger and Spotts compare the use of humor in television ads in the United States and in the United Kingdom. Suppose that independent random samples of television ads are taken in the two countries. A random sample of 400 television ads in the United Kingdom reveals that 142 use humor, while a random sample of 500 television ads in the United States reveals that 122 use humor.

a Set up the null and alternative hypotheses needed to determine whether the proportion of ads using humor in the United Kingdom differs from the proportion of ads using humor in the United States.

b Test the hypotheses you set up in part *a* by using critical values and by setting α equal to .10, .05, .01, and .001. How much evidence is there that the proportions of U.K. and U.S. ads using humor are different?

c Set up the hypotheses needed to attempt to establish that the difference between the proportions of U.K. and U.S. ads using humor is more than .05 (five percentage points). Test these hypotheses by using a p-value and by setting α equal to .10, .05, .01, and .001. How much evidence is there that the difference between the proportions exceeds .05?

d Calculate a 95 percent confidence interval for the difference between the proportion of U.K. ads using humor and the proportion of U.S. ads using humor. Interpret this interval. Can we be 95 percent confident that the proportion of U.K. ads using humor is greater than the proportion of U.S. ads using humor?

10.28 In the book *Essentials of Marketing Research,* William R. Dillon, Thomas J. Madden, and Neil H. Firtle discuss a research proposal in which a telephone company wants to determine whether the appeal of a new security system varies between homeowners and renters. Independent samples of 140 homeowners and 60 renters are randomly selected. Each respondent views a TV pilot in which a test ad for the new security system is embedded twice. Afterward, each respondent is interviewed to find out whether he or she would purchase the security system.

Results show that 25 out of the 140 homeowners definitely would buy the security system, while 9 out of the 60 renters definitely would buy the system.

a Letting p_1 be the proportion of homeowners who would buy the security system, and letting p_2 be the proportion of renters who would buy the security system, set up the null and alternative hypotheses needed to determine whether the proportion of homeowners who would buy the security system differs from the proportion of renters who would buy the security system.

b Find the test statistic z and the p-value for testing the hypotheses of part *a*. Use the p-value to test the hypotheses with α equal to .10, .05, .01, and .001. How much evidence is there that the proportions of homeowners and renters differ?

c Calculate a 95 percent confidence interval for the difference between the proportions of homeowners and renters who would buy the security system. On the basis of this interval, can we be 95 percent confident that these proportions differ? Explain.

Note: An Excel add-in (MegaStat) output of the hypothesis test and confidence interval in parts *b* and *c* is given in Appendix 10.2 on page 409.

10.29 In the book *Cases in Finance,* Nunnally and Plath present a case in which the estimated percentage of uncollectible accounts varies with the age of the account. Here the age of an unpaid account is the number of days elapsed since the invoice date.

An accountant believes that the percentage of accounts that will be uncollectible increases as the ages of the accounts increase. To test this theory, the accountant randomly selects independent samples of 500 accounts with ages between 31 and 60 days and 500 accounts with ages between 61 and 90 days from the accounts receivable ledger dated one year ago. When the sampled accounts are examined, it is found that 10 of the 500 accounts with ages between 31 and 60 days were eventually classified as uncollectible, while 27 of the 500 accounts with ages between 61 and 90 days were eventually classified as uncollectible. Let p_1 be the proportion of accounts with ages between 31 and 60 days that will be uncollectible, and let p_2 be the proportion of accounts with ages between 61 and 90 days that will be uncollectible.

a Use the MINITAB output below to determine how much evidence there is that we should reject H_0: $p_1 - p_2 = 0$ in favor of H_a: $p_1 - p_2 \neq 0$.

b Identify a 95 percent confidence interval for $p_1 - p_2$, and estimate the smallest that the difference between p_1 and p_2 might be.

Test and CI for Two Proportions

```
Sample                X     N    Sample p
1 (31 to 60 days)    10   500    0.020000        Difference = p(1) - p(2)
2 (61 to 90 days     27   500    0.054000        Estimate for difference:   -0.034

95% CI for difference:  (-0.0573036, -0.0106964)
Test for difference = 0 (vs not = 0):  Z = -2.85   P-Value = 0.004
```

10.30 On January 7, 2000, the Gallup Organization released the results of a poll comparing the lifestyles of today with yesteryear. The survey results were based on telephone interviews with a randomly selected national sample of 1,031 adults, 18 years and older, conducted December 20–21, 1999. The poll asked several questions and compared the 1999 responses with the responses given in polls taken in previous years. Below we summarize some of the poll's results.[6]

Percentage of respondents who

1	Had taken a vacation lasting six days or more within the last 12 months:	December 1999 42%	December 1968 62%
2	Took part in some sort of daily activity to keep physically fit:	December 1999 60%	September 1977 48%
3	Watched TV more than four hours on an average weekday:	December 1999 28%	April 1981 25%
4	Drove a car or truck to work:	December 1999 87%	April 1971 81%

Assuming that each poll was based on a randomly selected national sample of 1,031 adults and that the samples in different years are independent:

a Let p_1 be the December 1999 population proportion of U.S. adults who had taken a vacation lasting six days or more within the last 12 months, and let p_2 be the December 1968 population proportion who had taken such a vacation. Calculate a 99 percent confidence interval for the difference between p_1 and p_2. Interpret what this interval says about how these population proportions differ.

b Let p_1 be the December 1999 population proportion of U.S. adults who took part in some sort of daily activity to keep physically fit, and let p_2 be the September 1977 population proportion who did the same. Carry out a hypothesis test to attempt to justify that the proportion who took part in such daily activity increased from September 1977 to December 1999. Use $\alpha = .05$ and explain your result.

c Let p_1 be the December 1999 population proportion of U.S. adults who watched TV more than four hours on an average weekday, and let p_2 be the April 1981 population proportion who did the same. Carry out a hypothesis test to determine whether these population proportions differ. Use $\alpha = .05$ and interpret the result of your test.

d Let p_1 be the December 1999 population proportion of U.S. adults who drove a car or truck to work, and let p_2 be the April 1971 population proportion who did the same. Calculate a 95 percent confidence interval for the difference between p_1 and p_2. On the basis of this interval, can it be concluded that the 1999 and 1971 population proportions differ?

[6]Source: www.gallup.com/ The Gallup Poll, December 30, 1999.

Chapter Summary

This chapter has explained **how to compare two populations** by using confidence intervals and hypothesis tests. First we discussed how to compare **two population means** by using **independent samples.** Here the measurements in one sample are not related to the measurements in the other sample. When the population variances are unknown, ***t*-based** inferences are appropriate if the populations are normally distributed or the sample sizes are large. Both **equal variances and unequal variances *t*-based procedures** exist. We learned that, because it can be difficult to compare the population variances, many statisticians believe that it is almost always best to use the unequal variances procedure.

Sometimes samples are not independent. We learned that one such case is what is called a **paired difference experiment.** Here we obtain two different measurements on the same sample units, and we can compare two population means by using a confidence interval or by conducting a hypothesis test that employs the differences between the pairs of measurements. We concluded this chapter by discussing how to compare **two population proportions** by using **large, independent samples.**

Glossary of Terms

independent samples experiment: An experiment in which there is no relationship between the measurements in the different samples. (page 382)

paired difference experiment: An experiment in which two different measurements are taken on the same units and inferences are made using the differences between the pairs of measurements. (page 395)

sampling distribution of $\hat{p}_1 - \hat{p}_2$: The probability distribution that describes the population of all possible values of $\hat{p}_1 - \hat{p}_2$, where $\hat{p}_1$ is the sample proportion for a random sample taken from one population and $\hat{p}_2$ is the sample proportion for a random sample taken from a second population. (page 398)

sampling distribution of $\bar{x}_1 - \bar{x}_2$: The probability distribution that describes the population of all possible values of $\bar{x}_1 - \bar{x}_2$, where $\bar{x}_1$ is the sample mean of a random sample taken from one population and $\bar{x}_2$ is the sample mean of a random sample taken from a second population. (page 382)

Important Formulas and Tests

Sampling distribution of $\bar{x}_1 - \bar{x}_2$ (independent random samples): page 382

t-based confidence interval for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$: page 383

t-based confidence interval for $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$: page 386

t-test about $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$: page 385

t-test about $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$: page 386

Confidence interval for μ_d: page 393

A hypothesis test about μ_d: page 393

Sampling distribution of $\hat{p}_1 - \hat{p}_2$ (independent random samples): page 398

Large sample confidence interval for $p_1 - p_2$: page 399

Large sample hypothesis test about $p_1 - p_2$: page 400

Supplementary Exercises

connect

Exercises 10.31 and 10.32 deal with the following situation:

In an article in the *Journal of Retailing,* Kumar, Kerwin, and Pereira study factors affecting merger and acquisition activity in retailing by comparing "target firms" and "bidder firms" with respect to several financial and marketing-related variables. If we consider two of the financial variables included in the study, suppose a random sample of 36 "target firms" gives a mean earnings per share of \$1.52 with a standard deviation of \$0.92, and that this sample gives a mean debt-to-equity ratio of 1.66 with a standard deviation of 0.82. Furthermore, an independent random sample of 36 "bidder firms" gives a mean earnings per share of \$1.20 with a standard deviation of \$0.84, and this sample gives a mean debt-to-equity ratio of 1.58 with a standard deviation of 0.81.

10.31 **a** Set up the null and alternative hypotheses needed to test whether the mean earnings per share for all "target firms" differs from the mean earnings per share for all "bidder firms." Test these hypotheses at the .10, .05, .01, and .001 levels of significance. How much evidence is there that these means differ? Explain.

b Calculate a 95 percent confidence interval for the difference between the mean earnings per share for "target firms" and "bidder firms." Interpret the interval.

10.32 **a** Set up the null and alternative hypotheses needed to test whether the mean debt-to-equity ratio for all “target firms” differs from the mean debt-to-equity ratio for all “bidder firms.” Test these hypotheses at the .10, .05, .01, and .001 levels of significance. How much evidence is there that these means differ? Explain.

b Calculate a 95 percent confidence interval for the difference between the mean debt-to-equity ratios for “target firms” and “bidder firms.” Interpret the interval.

c Based on the results of this exercise and Exercise 10.31, does a firm’s earnings per share or the firm’s debt-to-equity ratio seem to have the most influence on whether a firm will be a “target” or a “bidder”? Explain.

10.33 What impact did the September 11 terrorist attack have on U.S. airline demand? An analysis was conducted by Ito and Lee, “Assessing the impact of the September 11 terrorist attacks on U.S. airline demand,” in the *Journal of Economics and Business* (January–February 2005). They found a negative short-term effect of over 30 percent and an ongoing negative impact of over 7 percent. Suppose that we wish to test the impact by taking a random sample of 12 airline routes before and after 9/11. Passenger miles (millions of passenger miles) for the same routes were tracked for the 12 months prior to and the 12 months immediately following 9/11. Assume that the population of all possible paired differences is normally distributed.

a Set up the null and alternative hypotheses needed to determine whether there was a reduction in mean airline passenger demand.

b Below we present the MINITAB output for the paired differences test. Use the output and critical values to test the hypotheses at the .10, .05, and .01 levels of significance. Has the population mean airline demand been reduced?

Paired T-Test and CI: Before911, After911

```
Paired T for Before911 - After911
              N      Mean     StDev  SE Mean
Before911    12   117.333    26.976    7.787
After911     12    87.583    25.518    7.366
Difference   12   29.7500   10.3056   2.9750

T-Test of mean difference = 0 (vs > 0): T-Value = 10.00  P-Value = 0.000
```

c Use the *p*-value to test the hypotheses at the .10, .05, and .01 levels of significance. How much evidence is there against the null hypothesis?

10.34 In the book *Essentials of Marketing Research,* William R. Dillon, Thomas J. Madden, and Neil H. Firtle discuss evaluating the effectiveness of a test coupon. Samples of 500 test coupons and 500 control coupons were randomly delivered to shoppers. The results indicated that 35 of the 500 control coupons were redeemed, while 50 of the 500 test coupons were redeemed.

a In order to consider the test coupon for use, the marketing research organization required that the proportion of all shoppers who would redeem the test coupon be statistically shown to be greater than the proportion of all shoppers who would redeem the control coupon. Assuming that the two samples of shoppers are independent, carry out a hypothesis test at the .01 level of significance that will show whether this requirement is met by the test coupon. Explain your conclusion.

b Use the sample data to find a point estimate and a 95 percent interval estimate of the difference between the proportions of all shoppers who would redeem the test coupon and the control coupon. What does this interval say about whether the test coupon should be considered for use? Explain.

c Carry out the test of part *a* at the .10 level of significance. What do you conclude? Is your result statistically significant? Compute a 90 percent interval estimate instead of the 95 percent interval estimate of part *b*. Based on the interval estimate, do you feel that this result is practically important? Explain.

10.35 A marketing manager wishes to compare the mean prices charged for two brands of CD players. The manager conducts a random survey of retail outlets and obtains independent random samples of prices with the following results:

	Onkyo	JVC
Sample mean, $\bar{x}$	\$189	\$145
Sample standard deviation, s	\$ 12	\$ 10
Sample size	6	12

Assuming normality and equal variances:

a Use an appropriate hypothesis test to determine whether the mean prices for the two brands differ. How much evidence is there that the mean prices differ?

b Use an appropriate 95 percent confidence interval to estimate the difference between the mean prices of the two brands of CD players. Do you think that the difference has practical importance?

c Use an appropriate hypothesis test to provide evidence supporting the claim that the mean price of the Onkyo CD player is more than \$30 higher than the mean price for the JVC CD player. Set α equal to .05.

10.36 In its February 2, 1998, issue, *Fortune* magazine published the results of a Yankelovich Partners survey of 600 adults that investigated their ideas about marriage and divorce. (All respondents had incomes of \$50,000 or more.). For each statement below, the proportions of men and women who agreed with the statement are given.

People were magnanimous on the general proposition:

- In a divorce in a long-term marriage where the husband works outside the home and the wife is not employed for pay, the wife should be entitled to half the assets accumulated during the marriage.
  93% of women agree
  85% of men agree

But when we got to the goodies, a gender gap began to appear . . .

- The pension accumulated during the marriage should be split evenly.
  80% of women agree
  68% of men agree
- Stock options granted during the marriage should be split evenly.
  77% of women agree
  62% of men agree

Source: Reprinted from the February 2, 1998, issue of *Fortune.* Copyright 1998 Time, Inc. Reprinted by permission.

Assuming that the survey results were obtained from independent random samples of 300 men and 300 women:

a For each statement, carry out a hypothesis test that tests the equality of the population proportions of men and women who agree with the statement. Use α equal to .10, .05, .01, and .001. How much evidence is there that the population proportions of men and women who agree with each statement differ?

b For each statement, calculate a 95 percent confidence interval for the difference between the population proportion of men who agree with the statement and the population proportion of women who agree with the statement. Use the interval to help assess whether you feel that the difference between population proportions has practical importance.

10.37 Internet Exercise

a A prominent issue of the 2000 U.S. presidential campaign was campaign finance reform. A *Washington Post*/ABC News poll (reported April 4, 2000) found that 63 percent of 1,083 American adults surveyed believed that stricter campaign finance laws would be effective (a lot or somewhat) in reducing the influence of money in politics. Was this view uniformly held or did it vary by gender, race, or political party affiliation? A summary of survey responses, broken down by gender, is given in the table below.

Summary of Responses	Male	Female	All
Believe reduce influence, p	59%	66%	63%
Number surveyed, n	520	563	1,083

[Source: *Washington Post* website: www.washingtonpost.com/wp-srv/politics/polls/vault/vault.htm.

Is there sufficient evidence in this survey to conclude that the proportion of individuals who believed that campaign finance laws can reduce the influence of money in politics differs between females and males? Set up the appropriate null and alternative hypotheses. Conduct your test at the .05 and .01 levels of significance and calculate the p-value for your test. Make sure your conclusion is clearly stated.

b Search the World Wide Web for an interesting recent political poll dealing with an issue or political candidates, where responses are broken down by gender or some other two-category classification. (A list of high-potential websites is given below.) Use a difference in proportions test to determine whether political preference differs by gender or other two-level grouping.

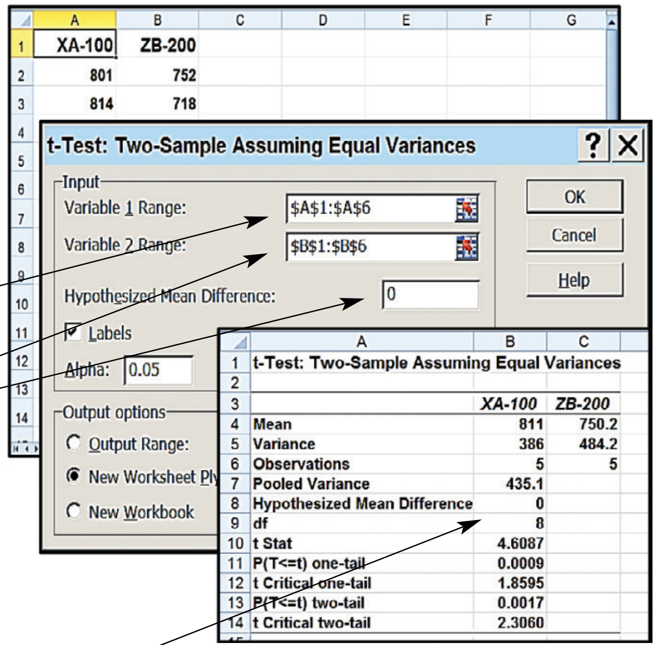
Political polls on the World Wide Web:

ABC News:	www.abcnews.go.com/pollingunit
Washington Post:	www.washingtonpost.com/wp-dyn/content/politics/polls/?nid=roll_polls
Gallup:	www.gallup.com/Home.aspx
Polling Report:	www.pollingreport.com
Rasmussen Reports:	www.rasmussenreports.com/public_content/politics
Zogby International:	www.zogby.com/features/zogbytables3.cfm
CBS News Poll Database:	www.cbsnews.com/stories/2007/10/12/politics/main3362530.shtml?tag=cbsnewsMainColumnArea;cbsnewsMainColumnArea.0

Appendix 10.1 ■ Two-Sample Hypothesis Testing Using Excel

Test for the difference between means, equal variances, in Figure 10.2(a) on page 386 (data file: Catalyst.xlsx):

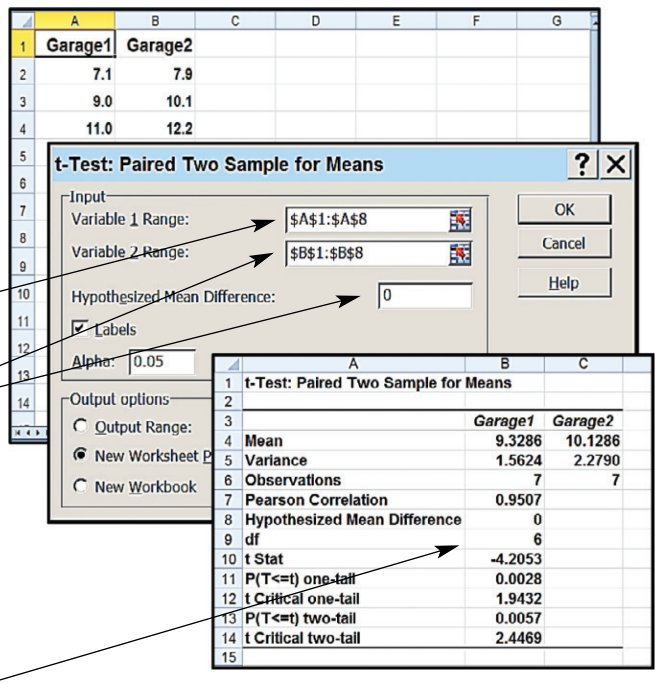
- Enter the data from Table 10.1 (page 384) into two columns: yields for catalyst XA-100 in column A and yields for catalyst ZB-200 in column B, with labels XA-100 and ZB-200.
- Select **Data : Data Analysis : t-Test: Two-Sample Assuming Equal Variances** and click OK in the Data Analysis dialog box.
- In the t-Test dialog box, enter A1: A6 in the "Variable 1 Range" window.
- Enter B1:B6 in the "Variable 2 Range" window.
- Enter 0 (zero) in the "Hypothesized Mean Difference" box.
- Place a checkmark in the Labels checkbox.
- Enter 0.05 into the Alpha box.
- Under output options, select "New Worksheet Ply" to have the output placed in a new worksheet and enter the name Output for the new worksheet.
- Click OK in the t-Test dialog box.
- The output will be displayed in a new worksheet.



Note: The *t*-test assuming unequal variances can be done by selecting **Data : Data Analysis : *t*-Test : Two-Sample Assuming Unequal Variances.**

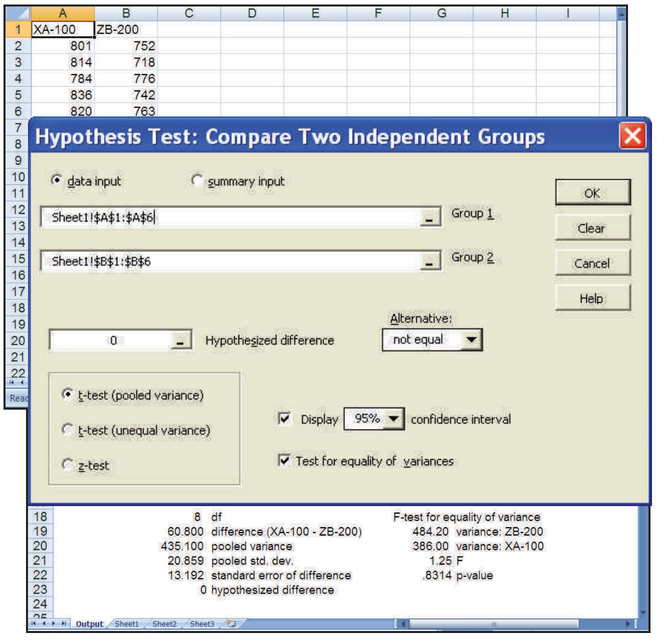
Test for paired differences in Figure 10.9 on page 394 (data file: Repair.xlsx):

- Enter the data from Table 10.2 (page 392) into two columns: costs for Garage 1 in column A and costs for Garage 2 in column B, with labels Garage1 and Garage2.
- Select **Data : Data Analysis : t-Test: Paired Two Sample for Means** and click OK in the Data Analysis dialog box.
- In the t-Test dialog box, enter A1:A8 into the "Variable 1 Range" window.
- Enter B1:B8 into the "Variable 2 Range" window.
- Enter 0 (zero) in the "Hypothesized Mean Difference" box.
- Place a checkmark in the Labels checkbox.
- Enter 0.05 into the Alpha box.
- Under output options, select "New Worksheet Ply" to have the output placed in a new worksheet and enter the name Output for the new worksheet.
- Click OK in the t-Test dialog box.
- The output will be displayed in a new worksheet.



Appendix 10.2 ■ Two-Sample Hypothesis Testing Using MegaStat

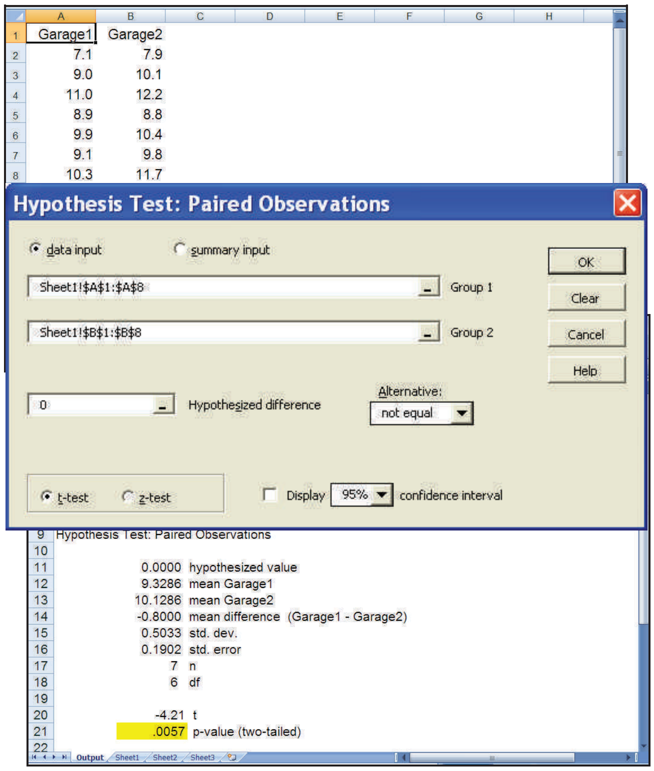
Test for the difference between means, equal variances, similar to Figure 10.2(a) on page 386 (data file: Catalyst.xlsx):



- Enter the data from Table 10.1 (page 384) into two columns: yields for catalyst XA-100 in column A and yields for catalyst ZB-200 in column B, with labels XA-100 and ZB-200.
- Select **MegaStat : Hypothesis Tests : Compare Two Independent Groups**
- In the "Hypothesis Test: Compare Two Independent Groups" dialog box, click on "data input."
- Click in the Group 1 window and use the autoexpand feature to enter the range A1:A6.
- Click in the Group 2 window and use the AutoExpand feature to enter the range B1:B6.
- Enter the Hypothesized Difference (here equal to 0) into the so labeled window.
- Select an Alternative (here "not equal") from the drop-down menu in the Alternative box.
- Click on "t-test (pooled variance)" to request the equal variances test described on page 385.
- Check the "Display confidence interval" checkbox, and select or type a desired level of confidence.
- Check the "Test for equality of variances" checkbox to request the *F*-test that will be discussed in Chapter 11.
- Click OK in the "Hypothesis Test: Compare Two Independent Groups" dialog box.
- The *t*-test assuming unequal variances described on page 386 can be done by clicking "t-test (unequal variance)."

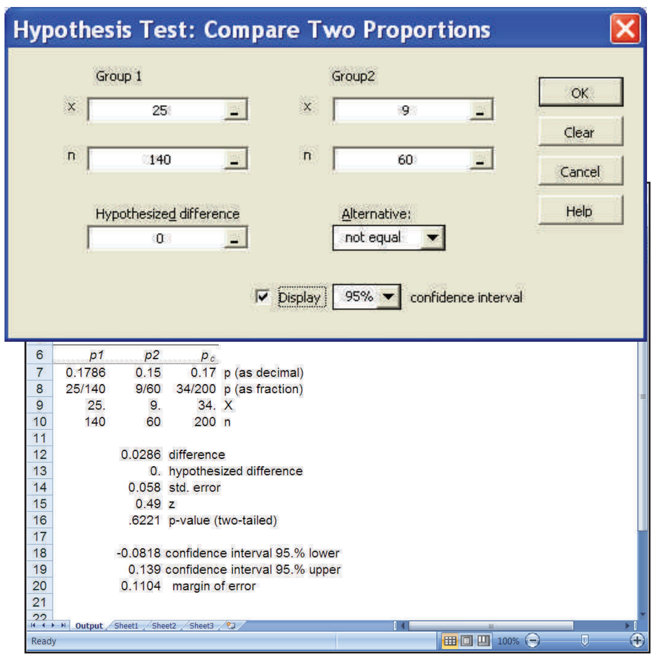
Test for paired differences similar to Figure 10.9 on page 394 (data file: Repair.xlsx):

- Enter the data from Table 10.2 (page 392) into two columns: costs for Garage 1 in column A and costs for Garage 2 in column B, with labels Garage1 and Garage2.
- Select **Add-Ins : MegaStat : Hypothesis Tests : Paired Observations.**
- In the "Hypothesis Test: Paired Observations" dialog box, click on "data input."
- Click in the Group 1 window, and use the AutoExpand feature to enter the range A1:A8.
- Click in the Group 2 window, and use the AutoExpand feature to enter the range B1:B8.
- Enter the Hypothesized difference (here equal to 0) into the so labeled window.
- Select an Alternative (here "not equal") from the drop-down menu in the Alternative box.
- Click on "t-test."
- Click OK in the "Hypothesis Test: Paired Observations" dialog box.
- If the sample sizes are large, a test based on the normal distribution can be done by clicking on "z-test."



Hypothesis Test and Confidence Interval for Two Independent Proportions in Exercise 10.28 on page 402:

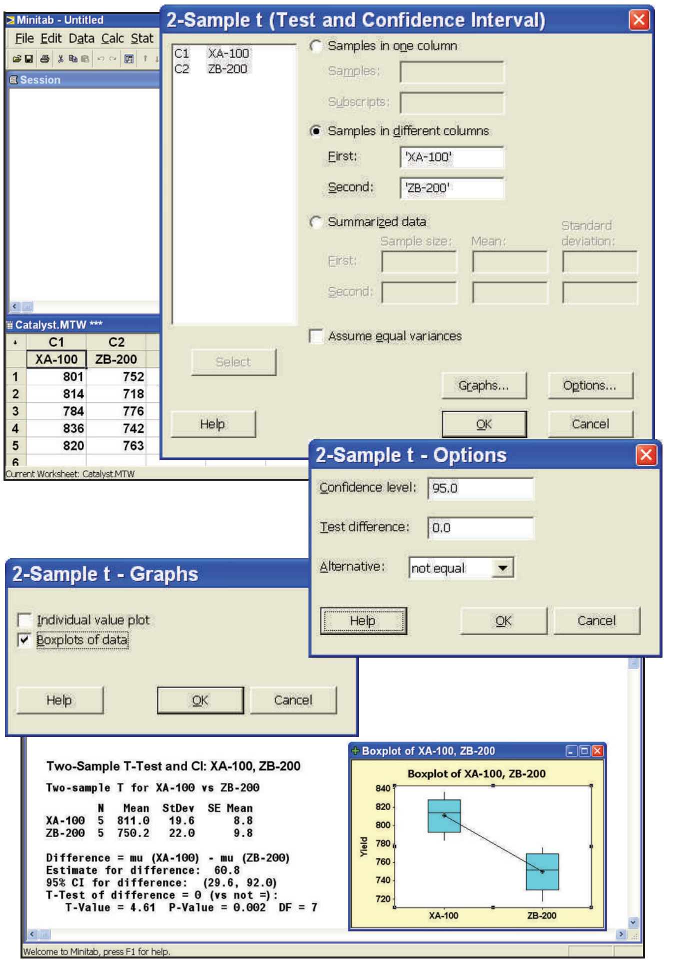
- Select **Add-Ins : MegaStat : Hypothesis Tests: Compare Two Independent Proportions.**
- In the "Hypothesis Test: Compare Two Proportions" dialog box, enter the number of successes *x* (here equal to 25) and the sample size *n* (here equal to 140) for homeowners in the "x" and "n" Group 1 windows.
- Enter the number of successes *x* (here equal to 9) and the sample size *n* (here equal to 60) for renters in the "x" and "n" Group 2 windows.
- Enter the Hypothesized difference (here equal to 0) into the so labeled window.
- Select an Alternative (here "not equal") from the drop-down menu in the Alternative box.
- Check the "Display confidence interval" checkbox, and select or type a desired level of confidence (here equal to 95%).
- Click OK in the "Hypothesis Test: Compare Two Proportions" dialog box.



Appendix 10.3 ■ Two-Sample Hypothesis Testing Using MINITAB

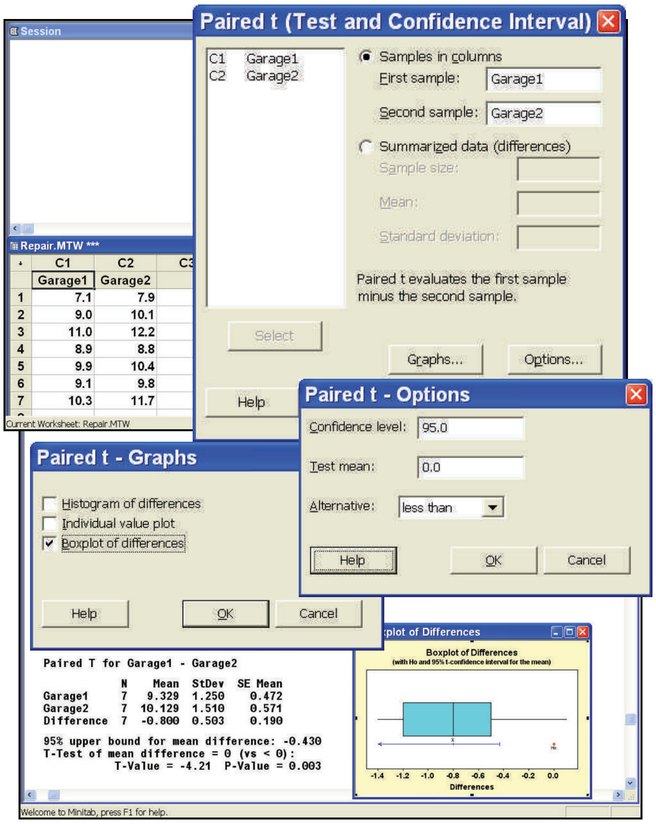
Test for the difference between means, unequal variances, in Figure 10.4 on page 388 (data file: Catalyst.MTW):

- In the data window, enter the data from Table 10.1 (page 384) into two columns with variable names XA-100 and ZB-200.
- Select **Stat : Basic Statistics : 2-Sample t**.
- In the “2-Sample t (Test and Confidence Interval)” dialog box, select the “Samples in different columns” option.
- Select the XA-100 variable into the First window.
- Select the ZB-200 variable into the Second window.
- Click on the Options . . . button, enter the desired level of confidence (here, 95.0) in the “Confidence level” window, enter 0.0 in the “Test difference” window, and select “not equal” from the Alternative pull-down menu. Click OK in the “2-Sample t—Options” dialog box.
- To produce yield by catalyst type boxplots, click the Graphs . . . button, check the “Boxplots of data” checkbox, and click OK in the “2 Sample t—Graphs” dialog box.
- Click OK in the “2-Sample t (Test and Confidence Interval)” dialog box.
- The results of the two-sample *t*-test (including the *t* statistic and *p*-value) and the confidence interval for the difference between means appear in the Session window, while the boxplots will be displayed in a graphics window.
- A test for the difference between two means when the **variances are equal** can be performed by placing a checkmark in the “Assume Equal Variances” checkbox in the “2-Sample t (Test and Confidence Interval)” dialog box.



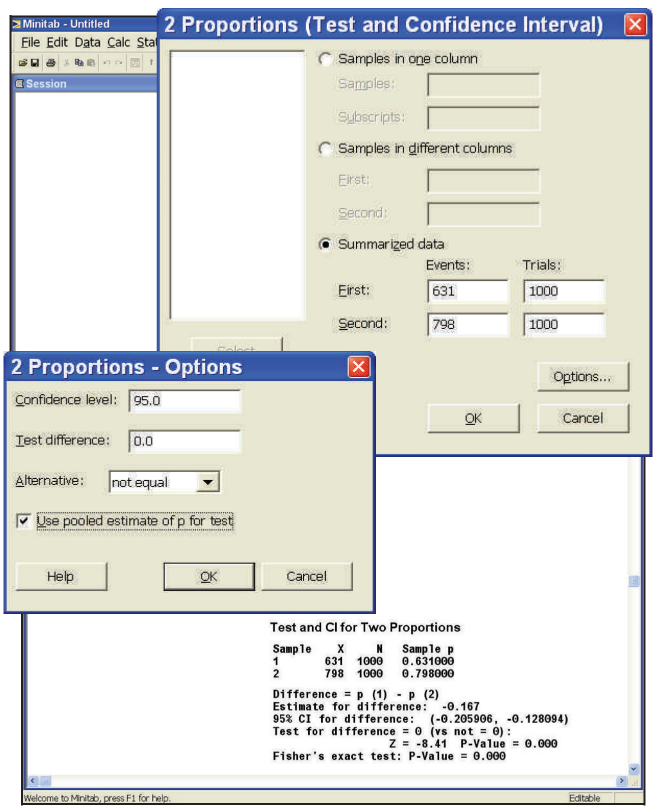
Test for paired differences in Figure 10.8 on page 394 (data file: Repair.MTW):

- In the Data window, enter the data from Table 10.2 (page 392) into two columns with variable names Garage1 and Garage2.
- Select **Stat : Basic Statistics : Paired t.**
- In the "Paired t (Test and Confidence Interval)" dialog box, select the "Samples in columns" option.
- Select Garage1 into the "First sample" window and Garage2 into the "Second sample" window.
- Click the Options . . . button.
- In the "Paired t—Options" dialog box, enter the desired level of confidence (here, 95.0) in the "Confidence level" window, enter 0.0 in the "Test mean" window, select "less than" from the Alternative pull-down menu, and click OK.
- To produce a boxplot of differences with a graphical summary of the test, click the Graphs . . . button, check the "Boxplot of differences" checkbox, and click OK in the "Paired t—Graphs" dialog box.
- Click OK in the "Paired t (Test and Confidence Interval)" dialog box. The results of the paired *t*-test are given in the Session window, and graphical output is displayed in a graphics window.



Hypothesis test and confidence interval for two Independent proportions in Figure 10.11 on page 401:

- Select **Stat : Basic Statistics : 2 Proportions.**
- In the "2 Proportions (Test and Confidence Interval)" dialog box, select the "Summarized data" option.
- Enter the sample size for Des Moines (equal to 1000) into the "First—Trials" window, and enter the number of successes for Des Moines (equal to 631) into the "First—Events" window.
- Enter the sample size for Toledo (equal to 1000) into the "Second—Trials" window, and enter the number of successes for Toledo (equal to 798) into the "Second—Events" window.
- Click on the Options . . . button.
- In the "2 Proportions—Options" dialog box, enter the desired level of confidence (here 95.0) into the "Confidence level" window.
- Enter 0.0 into the "Test difference" window because we are testing that the difference between the two proportions equals zero.
- Select the desired alternative hypothesis (here "not equal") from the Alternative drop-down menu.
- Check the "Use pooled estimate of p for test" checkbox because "Test difference" equals zero. Do not check this box in cases where "Test difference" does not equal zero.
- Click OK in the "2 Proportions—Options" dialog box.



- Click OK in the "2 Proportions (Test and Confidence Interval)" dialog box to obtain results for the test in the Session window.