



# CHAPTER 10

## Inference About Means and Proportions with Two Populations

---

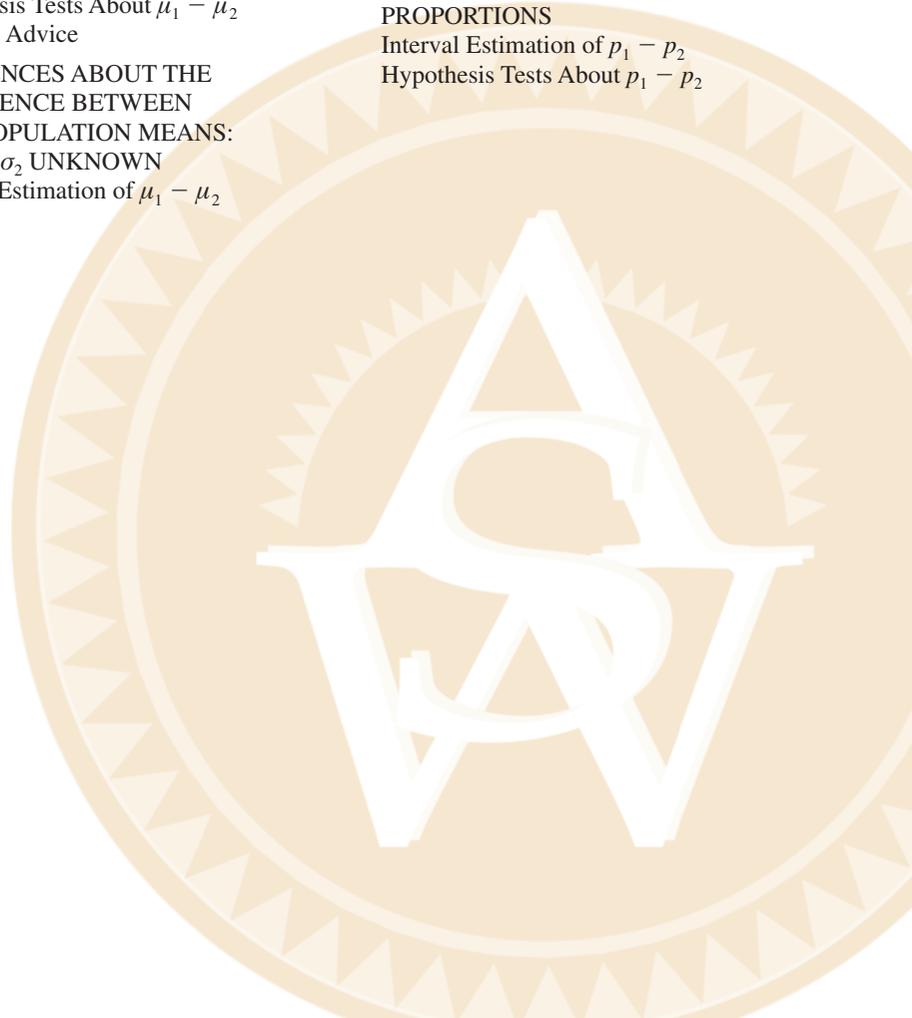
### CONTENTS

STATISTICS IN PRACTICE: U.S. FOOD AND DRUG ADMINISTRATION

- 10.1** INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS:  $\sigma_1$  AND  $\sigma_2$  KNOWN  
Interval Estimation of  $\mu_1 - \mu_2$   
Hypothesis Tests About  $\mu_1 - \mu_2$   
Practical Advice
- 10.2** INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS:  $\sigma_1$  AND  $\sigma_2$  UNKNOWN  
Interval Estimation of  $\mu_1 - \mu_2$

Hypothesis Tests About  $\mu_1 - \mu_2$   
Practical Advice

- 10.3** INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION MEANS: MATCHED SAMPLES
- 10.4** INFERENCES ABOUT THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS  
Interval Estimation of  $p_1 - p_2$   
Hypothesis Tests About  $p_1 - p_2$



**STATISTICS** *in* **PRACTICE****U.S. FOOD AND DRUG ADMINISTRATION**  
*WASHINGTON, D.C.*

It is the responsibility of the U.S. Food and Drug Administration (FDA), through its Center for Drug Evaluation and Research (CDER), to ensure that drugs are safe and effective. But CDER does not do the actual testing of new drugs itself. It is the responsibility of the company seeking to market a new drug to test it and submit evidence that it is safe and effective. CDER statisticians and scientists then review the evidence submitted.

Companies seeking approval of a new drug conduct extensive statistical studies to support their application. The testing process in the pharmaceutical industry usually consists of three stages: (1) preclinical testing, (2) testing for long-term usage and safety, and (3) clinical efficacy testing. At each successive stage, the chance that a drug will pass the rigorous tests decreases; however, the cost of further testing increases dramatically. Industry surveys indicate that on average the research and development for one new drug costs \$250 million and takes 12 years. Hence, it is important to eliminate unsuccessful new drugs in the early stages of the testing process, as well as to identify promising ones for further testing.

Statistics plays a major role in pharmaceutical research, where government regulations are stringent and rigorously enforced. In preclinical testing, a two- or three-population statistical study typically is used to determine whether a new drug should continue to be studied in the long-term usage and safety program. The populations may consist of the new drug, a control, and a standard drug. The preclinical testing process begins when a new drug is sent to the pharmacology group for evaluation of efficacy—the capacity of the drug to produce the desired effects. As part of the process, a statistician is asked to design an experiment that can be used to test the new drug. The design must specify the sample size and the statistical methods of analysis. In a two-population study, one sample is used to obtain data on the efficacy of the new drug (population 1) and a second sample is used to obtain data on the efficacy of a standard drug (population 2). Depending on the intended use, the new and standard drugs are tested in such disciplines as neurology, cardiology, and immunology. In most studies, the statistical method involves hypothesis testing for the difference between the means of the new drug population



Statistical methods are used to test and develop new drugs. © Lester Lefkowitz/CORBIS.

and the standard drug population. If a new drug lacks efficacy or produces undesirable effects in comparison with the standard drug, the new drug is rejected and withdrawn from further testing. Only new drugs that show promising comparisons with the standard drugs are forwarded to the long-term usage and safety testing program.

Further data collection and multipopulation studies are conducted in the long-term usage and safety testing program and in the clinical testing programs. The FDA requires that statistical methods be defined prior to such testing to avoid data-related biases. In addition, to avoid human biases, some of the clinical trials are double or triple blind. That is, neither the subject nor the investigator knows what drug is administered to whom. If the new drug meets all requirements in relation to the standard drug, a new drug application (NDA) is filed with the FDA. The application is rigorously scrutinized by statisticians and scientists at the agency.

In this chapter you will learn how to construct interval estimates and make hypothesis tests about means and proportions with two populations. Techniques will be presented for analyzing independent random samples as well as matched samples.

In Chapters 8 and 9 we showed how to develop interval estimates and conduct hypothesis tests for situations involving a single population mean and a single population proportion. In this chapter we continue our discussion of statistical inference by showing how interval estimates and hypothesis tests can be developed for situations involving two populations when the difference between the two population means or the two population proportions is of prime importance. For example, we may want to develop an interval estimate of the difference between the mean starting salary for a population of men and the mean starting salary for a population of women or conduct a hypothesis test to determine whether any difference is present between the proportion of defective parts in a population of parts produced by supplier A and the proportion of defective parts in a population of parts produced by supplier B. We begin our discussion of statistical inference about two populations by showing how to develop interval estimates and conduct hypothesis tests about the difference between the means of two populations when the standard deviations of the two populations are assumed known.

## 10.1

## Inferences About the Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Known

Letting  $\mu_1$  denote the mean of population 1 and  $\mu_2$  denote the mean of population 2, we will focus on inferences about the difference between the means:  $\mu_1 - \mu_2$ . To make an inference about this difference, we select a simple random sample of  $n_1$  units from population 1 and a second simple random sample of  $n_2$  units from population 2. The two samples, taken separately and independently, are referred to as **independent simple random samples**. In this section, we assume that information is available such that the two population standard deviations,  $\sigma_1$  and  $\sigma_2$ , can be assumed known prior to collecting the samples. We refer to this situation as the  $\sigma_1$  and  $\sigma_2$  known case. In the following example we show how to compute a margin of error and develop an interval estimate of the difference between the two population means when  $\sigma_1$  and  $\sigma_2$  are known.

### Interval Estimation of $\mu_1 - \mu_2$

Greystone Department Stores, Inc., operates two stores in Buffalo, New York: One is in the inner city and the other is in a suburban shopping center. The regional manager noticed that products that sell well in one store do not always sell well in the other. The manager believes this situation may be attributable to differences in customer demographics at the two locations. Customers may differ in age, education, income, and so on. Suppose the manager asks us to investigate the difference between the mean ages of the customers who shop at the two stores.

Let us define population 1 as all customers who shop at the inner-city store and population 2 as all customers who shop at the suburban store.

$\mu_1$  = mean of population 1 (i.e., the mean age of all customers who shop at the inner-city store)

$\mu_2$  = mean of population 2 (i.e., the mean age of all customers who shop at the suburban store)

The difference between the two population means is  $\mu_1 - \mu_2$ .

To estimate  $\mu_1 - \mu_2$ , we will select a simple random sample of  $n_1$  customers from population 1 and a simple random sample of  $n_2$  customers from population 2. We then compute the two sample means.

$\bar{x}_1$  = sample mean age for the simple random sample of  $n_1$  inner-city customers

$\bar{x}_2$  = sample mean age for the simple random sample of  $n_2$  suburban customers

The point estimator of the difference between the two population means is the difference between the two sample means.

POINT ESTIMATOR OF THE DIFFERENCE BETWEEN TWO POPULATION MEANS

$$\bar{x}_1 - \bar{x}_2 \quad (10.1)$$

Figure 10.1 provides an overview of the process used to estimate the difference between two population means based on two independent simple random samples.

As with other point estimators, the point estimator  $\bar{x}_1 - \bar{x}_2$  has a standard error that describes the variation in the sampling distribution of the estimator. With two independent simple random samples, the standard error of  $\bar{x}_1 - \bar{x}_2$  is as follows:

*The standard error of  $\bar{x}_1 - \bar{x}_2$  is the standard deviation of the sampling distribution of  $\bar{x}_1 - \bar{x}_2$ .*

STANDARD ERROR OF  $\bar{x}_1 - \bar{x}_2$

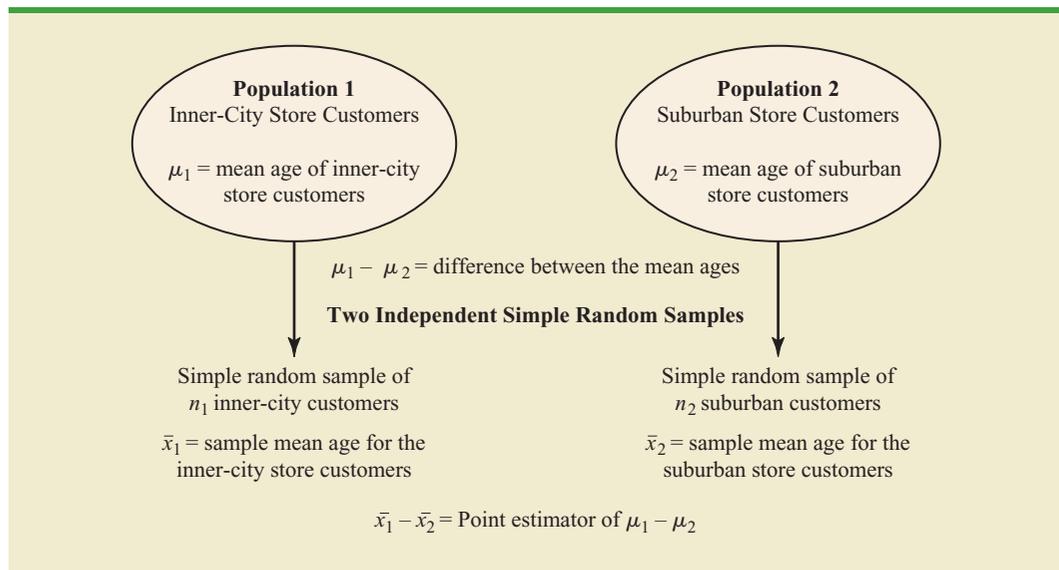
$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.2)$$

If both populations have a normal distribution, or if the sample sizes are large enough that the central limit theorem enables us to conclude that the sampling distributions of  $\bar{x}_1$  and  $\bar{x}_2$  can be approximated by a normal distribution, the sampling distribution of  $\bar{x}_1 - \bar{x}_2$  will have a normal distribution with mean given by  $\mu_1 - \mu_2$ .

As we showed in Chapter 8, an interval estimate is given by a point estimate  $\pm$  a margin of error. In the case of estimation of the difference between two population means, an interval estimate will take the following form:

$$\bar{x}_1 - \bar{x}_2 \pm \text{Margin of error}$$

**FIGURE 10.1** ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS



With the sampling distribution of  $\bar{x}_1 - \bar{x}_2$  having a normal distribution, we can write the margin of error as follows:

The margin of error is given by multiplying the standard error by  $z_{\alpha/2}$ .

$$\text{Margin of error} = z_{\alpha/2} \sigma_{\bar{x}_1 - \bar{x}_2} = z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.3)$$

Thus the interval estimate of the difference between two population means is as follows:

INTERVAL ESTIMATE OF THE DIFFERENCE BETWEEN TWO POPULATION MEANS:  $\sigma_1$  AND  $\sigma_2$  KNOWN

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.4)$$

where  $1 - \alpha$  is the confidence coefficient.

Let us return to the Greystone example. Based on data from previous customer demographic studies, the two population standard deviations are known with  $\sigma_1 = 9$  years and  $\sigma_2 = 10$  years. The data collected from the two independent simple random samples of Greystone customers provided the following results.

	Inner City Store	Suburban Store
Sample Size	$n_1 = 36$	$n_2 = 49$
Sample Mean	$\bar{x}_1 = 40$ years	$\bar{x}_2 = 35$ years

Using expression (10.1), we find that the point estimate of the difference between the mean ages of the two populations is  $\bar{x}_1 - \bar{x}_2 = 40 - 35 = 5$  years. Thus, we estimate that the customers at the inner-city store have a mean age five years greater than the mean age of the suburban store customers. We can now use expression (10.4) to compute the margin of error and provide the interval estimate of  $\mu_1 - \mu_2$ . Using 95% confidence and  $z_{\alpha/2} = z_{0.025} = 1.96$ , we have

$$\begin{aligned} \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\ 40 - 35 \pm 1.96 \sqrt{\frac{9^2}{36} + \frac{10^2}{49}} \\ 5 \pm 4.06 \end{aligned}$$

Thus, the margin of error is 4.06 years and the 95% confidence interval estimate of the difference between the two population means is  $5 - 4.06 = .94$  years to  $5 + 4.06 = 9.06$  years.

## Hypothesis Tests About $\mu_1 - \mu_2$

Let us consider hypothesis tests about the difference between two population means. Using  $D_0$  to denote the hypothesized difference between  $\mu_1$  and  $\mu_2$ , the three forms for a hypothesis test are as follows:

$$\begin{array}{lll} H_0: \mu_1 - \mu_2 \geq D_0 & H_0: \mu_1 - \mu_2 \leq D_0 & H_0: \mu_1 - \mu_2 = D_0 \\ H_a: \mu_1 - \mu_2 < D_0 & H_a: \mu_1 - \mu_2 > D_0 & H_a: \mu_1 - \mu_2 \neq D_0 \end{array}$$

In many applications,  $D_0 = 0$ . Using the two-tailed test as an example, when  $D_0 = 0$  the null hypothesis is  $H_0: \mu_1 - \mu_2 = 0$ . In this case, the null hypothesis is that  $\mu_1$  and  $\mu_2$  are equal. Rejection of  $H_0$  leads to the conclusion that  $H_a: \mu_1 - \mu_2 \neq 0$  is true; that is,  $\mu_1$  and  $\mu_2$  are not equal.

The steps for conducting hypothesis tests presented in Chapter 9 are applicable here. We must choose a level of significance, compute the value of the test statistic and find the  $p$ -value to determine whether the null hypothesis should be rejected. With two independent simple random samples, we showed that the point estimator  $\bar{x}_1 - \bar{x}_2$  has a standard error  $\sigma_{\bar{x}_1 - \bar{x}_2}$  given by expression (10.2) and, when the sample sizes are large enough, the distribution of  $\bar{x}_1 - \bar{x}_2$  can be described by a normal distribution. In this case, the test statistic for the difference between two population means when  $\sigma_1$  and  $\sigma_2$  are known is as follows.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT  $\mu_1 - \mu_2$ :  $\sigma_1$  AND  $\sigma_2$  KNOWN

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (10.5)$$

Let us demonstrate the use of this test statistic in the following hypothesis testing example.

As part of a study to evaluate differences in education quality between two training centers, a standardized examination is given to individuals who are trained at the centers. The difference between the mean examination scores is used to assess quality differences between the centers. The population means for the two centers are as follows.

$\mu_1$  = the mean examination score for the population  
of individuals trained at center A

$\mu_2$  = the mean examination score for the population  
of individuals trained at center B

We begin with the tentative assumption that no difference exists between the training quality provided at the two centers. Hence, in terms of the mean examination scores, the null hypothesis is that  $\mu_1 - \mu_2 = 0$ . If sample evidence leads to the rejection of this hypothesis, we will conclude that the mean examination scores differ for the two populations. This conclusion indicates a quality differential between the two centers and suggests that a follow-up study investigating the reason for the differential may be warranted. The null and alternative hypotheses for this two-tailed test are written as follows.

$$\begin{aligned} H_0: \mu_1 - \mu_2 &= 0 \\ H_a: \mu_1 - \mu_2 &\neq 0 \end{aligned}$$

The standardized examination given previously in a variety of settings always resulted in an examination score standard deviation near 10 points. Thus, we will use this information to assume that the population standard deviations are known with  $\sigma_1 = 10$  and  $\sigma_2 = 10$ . An  $\alpha = .05$  level of significance is specified for the study.

Independent simple random samples of  $n_1 = 30$  individuals from training center A and  $n_2 = 40$  individuals from training center B are taken. The respective sample means are  $\bar{x}_1 = 82$  and  $\bar{x}_2 = 78$ . Do these data suggest a significant difference between the population

means at the two training centers? To help answer this question, we compute the test statistic using equation (10.5).

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(82 - 78) - 0}{\sqrt{\frac{10^2}{30} + \frac{10^2}{40}}} = 1.66$$

Next let us compute the  $p$ -value for this two-tailed test. Because the test statistic  $z$  is in the upper tail, we first compute the area under the curve to the right of  $z = 1.66$ . Using the standard normal distribution table, the area to the left of  $z = 1.66$  is .9515. Thus, the area in the upper tail of the distribution is  $1.0000 - .9515 = .0485$ . Because this test is a two-tailed test, we must double the tail area:  $p\text{-value} = 2(.0485) = .0970$ . Following the usual rule to reject  $H_0$  if  $p\text{-value} \leq \alpha$ , we see that the  $p$ -value of .0970 does not allow us to reject  $H_0$  at the .05 level of significance. The sample results do not provide sufficient evidence to conclude the training centers differ in quality.

In this chapter we will use the  $p$ -value approach to hypothesis testing as described in Chapter 9. However, if you prefer, the test statistic and the critical value rejection rule may be used. With  $\alpha = .05$  and  $z_{\alpha/2} = z_{.025} = 1.96$ , the rejection rule employing the critical value approach would be reject  $H_0$  if  $z \leq -1.96$  or if  $z \geq 1.96$ . With  $z = 1.66$ , we reach the same do not reject  $H_0$  conclusion.

In the preceding example, we demonstrated a two-tailed hypothesis test about the difference between two population means. Lower tail and upper tail tests can also be considered. These tests use the same test statistic as given in equation (10.5). The procedure for computing the  $p$ -value and the rejection rules for these one-tailed tests are the same as those presented in Chapter 9.

## Practical Advice

In most applications of the interval estimation and hypothesis testing procedures presented in this section, random samples with  $n_1 \geq 30$  and  $n_2 \geq 30$  are adequate. In cases where either or both sample sizes are less than 30, the distributions of the populations become important considerations. In general, with smaller sample sizes, it is more important for the analyst to be satisfied that it is reasonable to assume that the distributions of the two populations are at least approximately normal.

## Exercises

### Methods

- The following results come from two independent random samples taken of two populations.

**SELF test**

Sample 1	Sample 2
$n_1 = 50$	$n_2 = 35$
$\bar{x}_1 = 13.6$	$\bar{x}_2 = 11.6$
$\sigma_1 = 2.2$	$\sigma_2 = 3.0$

- What is the point estimate of the difference between the two population means?
- Provide a 90% confidence interval for the difference between the two population means.
- Provide a 95% confidence interval for the difference between the two population means.

**SELF test**

2. Consider the following hypothesis test.

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_a: \mu_1 - \mu_2 > 0$$

The following results are for two independent samples taken from the two populations.

Sample 1	Sample 2
$n_1 = 40$	$n_2 = 50$
$\bar{x}_1 = 25.2$	$\bar{x}_2 = 22.8$
$\sigma_1 = 5.2$	$\sigma_2 = 6.0$

- What is the value of the test statistic?
  - What is the  $p$ -value?
  - With  $\alpha = .05$ , what is your hypothesis testing conclusion?
3. Consider the following hypothesis test.

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

$$H_a: \mu_1 - \mu_2 \neq 0$$

The following results are for two independent samples taken from the two populations.

Sample 1	Sample 2
$n_1 = 80$	$n_2 = 70$
$\bar{x}_1 = 104$	$\bar{x}_2 = 106$
$\sigma_1 = 8.4$	$\sigma_2 = 7.6$

- What is the value of the test statistic?
- What is the  $p$ -value?
- With  $\alpha = .05$ , what is your hypothesis testing conclusion?

**Applications****SELF test**

4. *Condé Nast Traveler* conducts an annual survey in which readers rate their favorite cruise ship. All ships are rated on a 100-point scale, with higher values indicating better service. A sample of 37 ships that carry fewer than 500 passengers resulted in an average rating of 85.36, and a sample of 44 ships that carry 500 or more passengers provided an average rating of 81.40 (*Condé Nast Traveler*; February 2008). Assume that the population standard deviation is 4.55 for ships that carry fewer than 500 passengers and 3.97 for ships that carry 500 or more passengers.
- What is the point estimate of the difference between the population mean rating for ships that carry fewer than 500 passengers and the population mean rating for ships that carry 500 or more passengers?
  - At 95% confidence, what is the margin of error?
  - What is a 95% confidence interval estimate of the difference between the population mean ratings for the two sizes of ships?
5. The average expenditure on Valentine's Day was expected to be \$100.89 (*USA Today*, February 13, 2006). Do male and female consumers differ in the amounts they spend? The average expenditure in a sample survey of 40 male consumers was \$135.67, and the average expenditure in a sample survey of 30 female consumers was \$68.64. Based on past surveys, the standard deviation for male consumers is assumed to be \$35, and the standard deviation for female consumers is assumed to be \$20.

**WEB file****Hotel**

- a. What is the point estimate of the difference between the population mean expenditure for males and the population mean expenditure for females?
  - b. At 99% confidence, what is the margin of error?
  - c. Develop a 99% confidence interval for the difference between the two population means.
6. Suppose that you are responsible for making arrangements for a business convention. Because of budget cuts due to the recent recession, you have been charged with choosing a city for the convention that has the least expensive hotel rooms. You have narrowed your choices to Atlanta and Houston. The file named *Hotel* contains samples of prices for rooms in Atlanta and Houston that are consistent with the results reported by Smith Travel Research (*SmartMoney*, March 2009). Because considerable historical data on the prices of rooms in both cities are available, the population standard deviations for the prices can be assumed to be \$20 in Atlanta and \$25 in Houston. Based on the sample data, can you conclude that the mean price of a hotel room in Atlanta is lower than one in Houston?
7. During the 2003 season, Major League Baseball took steps to speed up the play of baseball games in order to maintain fan interest (CNN Headline News, September 30, 2003). The following results come from a sample of 60 games played during the summer of 2002 and a sample of 50 games played during the summer of 2003. The sample mean shows the mean duration of the games included in each sample.

2002 Season	2003 Season
$n_1 = 60$	$n_2 = 50$
$\bar{x}_1 = 2 \text{ hours, } 52 \text{ minutes}$	$\bar{x}_2 = 2 \text{ hours, } 46 \text{ minutes}$

- a. A research hypothesis was that the steps taken during the 2003 season would reduce the population mean duration of baseball games. Formulate the null and alternative hypotheses.
  - b. What is the point estimate of the reduction in the mean duration of games during the 2003 season?
  - c. Historical data indicate a population standard deviation of 12 minutes is a reasonable assumption for both years. Conduct the hypothesis test and report the  $p$ -value. At a .05 level of significance, what is your conclusion?
  - d. Provide a 95% confidence interval estimate of the reduction in the mean duration of games during the 2003 season.
  - e. What was the percentage reduction in the mean time of baseball games during the 2003 season? Should management be pleased with the results of the statistical analysis? Discuss. Should the length of baseball games continue to be an issue in future years? Explain.
8. Will improving customer service result in higher stock prices for the companies providing the better service? “When a company’s satisfaction score has improved over the prior year’s results and is above the national average (currently 75.7), studies show its shares have a good chance of outperforming the broad stock market in the long run” (*BusinessWeek*, March 2, 2009). The following satisfaction scores of three companies for the 4th quarters of 2007 and 2008 were obtained from the American Customer Satisfaction Index. Assume that the scores are based on a poll of 60 customers from each company. Because the polling has been done for several years, the standard deviation can be assumed to equal 6 points in each case.

Company	2007 Score	2008 Score
Rite Aid	73	76
Expedia	75	77
J.C. Penney	77	78

- For Rite Aid, is the increase in the satisfaction score from 2007 to 2008 statistically significant? Use  $\alpha = .05$ . What can you conclude?
- Can you conclude that the 2008 score for Rite Aid is above the national average of 75.7? Use  $\alpha = .05$ .
- For Expedia, is the increase from 2007 to 2008 statistically significant? Use  $\alpha = .05$ .
- When conducting a hypothesis test with the values given for the standard deviation, sample size, and  $\alpha$ , how large must the increase from 2007 to 2008 be for it to be statistically significant?
- Use the result of part (d) to state whether the increase for J.C. Penney from 2007 to 2008 is statistically significant.

## 10.2

## Inferences About the Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Unknown

In this section we extend the discussion of inferences about the difference between two population means to the case when the two population standard deviations,  $\sigma_1$  and  $\sigma_2$ , are unknown. In this case, we will use the sample standard deviations,  $s_1$  and  $s_2$ , to estimate the unknown population standard deviations. When we use the sample standard deviations, the interval estimation and hypothesis testing procedures will be based on the  $t$  distribution rather than the standard normal distribution.

### Interval Estimation of $\mu_1 - \mu_2$

In the following example we show how to compute a margin of error and develop an interval estimate of the difference between two population means when  $\sigma_1$  and  $\sigma_2$  are unknown. Clearwater National Bank is conducting a study designed to identify differences between checking account practices by customers at two of its branch banks. A simple random sample of 28 checking accounts is selected from the Cherry Grove Branch and an independent simple random sample of 22 checking accounts is selected from the Beechmont Branch. The current checking account balance is recorded for each of the checking accounts. A summary of the account balances follows:

**WEB file**  
CheckAcct

	Cherry Grove	Beechmont
Sample Size	$n_1 = 28$	$n_2 = 22$
Sample Mean	$\bar{x}_1 = \$1025$	$\bar{x}_2 = \$910$
Sample Standard Deviation	$s_1 = \$150$	$s_2 = \$125$

Clearwater National Bank would like to estimate the difference between the mean checking account balance maintained by the population of Cherry Grove customers and the population of Beechmont customers. Let us develop the margin of error and an interval estimate of the difference between these two population means.

In Section 10.1, we provided the following interval estimate for the case when the population standard deviations,  $\sigma_1$  and  $\sigma_2$ , are known.

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

When  $\sigma_1$  and  $\sigma_2$  are estimated by  $s_1$  and  $s_2$ , the  $t$  distribution is used to make inferences about the difference between two population means.

With  $\sigma_1$  and  $\sigma_2$  unknown, we will use the sample standard deviations  $s_1$  and  $s_2$  to estimate  $\sigma_1$  and  $\sigma_2$  and replace  $z_{\alpha/2}$  with  $t_{\alpha/2}$ . As a result, the interval estimate of the difference between two population means is given by the following expression:

INTERVAL ESTIMATE OF THE DIFFERENCE BETWEEN TWO POPULATION MEANS:  $\sigma_1$  AND  $\sigma_2$  UNKNOWN

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (10.6)$$

where  $1 - \alpha$  is the confidence coefficient.

In this expression, the use of the  $t$  distribution is an approximation, but it provides excellent results and is relatively easy to use. The only difficulty that we encounter in using expression (10.6) is determining the appropriate degrees of freedom for  $t_{\alpha/2}$ . Statistical software packages compute the appropriate degrees of freedom automatically. The formula used is as follows:

DEGREES OF FREEDOM:  $t$  DISTRIBUTION WITH TWO INDEPENDENT RANDOM SAMPLES

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} \quad (10.7)$$

Let us return to the Clearwater National Bank example and show how to use expression (10.6) to provide a 95% confidence interval estimate of the difference between the population mean checking account balances at the two branch banks. The sample data show  $n_1 = 28$ ,  $\bar{x}_1 = \$1025$ , and  $s_1 = \$150$  for the Cherry Grove branch, and  $n_2 = 22$ ,  $\bar{x}_2 = \$910$ , and  $s_2 = \$125$  for the Beechmont branch. The calculation for degrees of freedom for  $t_{\alpha/2}$  is as follows:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{150^2}{28} + \frac{125^2}{22}\right)^2}{\frac{1}{28 - 1} \left(\frac{150^2}{28}\right)^2 + \frac{1}{22 - 1} \left(\frac{125^2}{22}\right)^2} = 47.8$$

We round the noninteger degrees of freedom down to 47 to provide a larger  $t$ -value and a more conservative interval estimate. Using the  $t$  distribution table with 47 degrees of freedom, we find  $t_{.025} = 2.012$ . Using expression (10.6), we develop the 95% confidence interval estimate of the difference between the two population means as follows.

$$\begin{aligned} \bar{x}_1 - \bar{x}_2 \pm t_{.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ 1025 - 910 \pm 2.012 \sqrt{\frac{150^2}{28} + \frac{125^2}{22}} \\ 115 \pm 78 \end{aligned}$$

The point estimate of the difference between the population mean checking account balances at the two branches is \$115. The margin of error is \$78, and the 95% confidence interval

estimate of the difference between the two population means is  $115 - 78 = \$37$  to  $115 + 78 = \$193$ .

*This suggestion should help if you are using equation (10.7) to calculate the degrees of freedom by hand.*

The computation of the degrees of freedom (equation (10.7)) is cumbersome if you are doing the calculation by hand, but it is easily implemented with a computer software package. However, note that the expressions  $s_1^2/n_1$  and  $s_2^2/n_2$  appear in both expression (10.6) and equation (10.7). These values only need to be computed once in order to evaluate both (10.6) and (10.7).

## Hypothesis Tests About $\mu_1 - \mu_2$

Let us now consider hypothesis tests about the difference between the means of two populations when the population standard deviations  $\sigma_1$  and  $\sigma_2$  are unknown. Letting  $D_0$  denote the hypothesized difference between  $\mu_1$  and  $\mu_2$ , Section 10.1 showed that the test statistic used for the case where  $\sigma_1$  and  $\sigma_2$  are known is as follows.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The test statistic,  $z$ , follows the standard normal distribution.

When  $\sigma_1$  and  $\sigma_2$  are unknown, we use  $s_1$  as an estimator of  $\sigma_1$  and  $s_2$  as an estimator of  $\sigma_2$ . Substituting these sample standard deviations for  $\sigma_1$  and  $\sigma_2$  provides the following test statistic when  $\sigma_1$  and  $\sigma_2$  are unknown.

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT  $\mu_1 - \mu_2$ :  $\sigma_1$  AND  $\sigma_2$  UNKNOWN

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (10.8)$$

The degrees of freedom for  $t$  are given by equation (10.7).

Let us demonstrate the use of this test statistic in the following hypothesis testing example.

Consider a new computer software package developed to help systems analysts reduce the time required to design, develop, and implement an information system. To evaluate the benefits of the new software package, a random sample of 24 systems analysts is selected. Each analyst is given specifications for a hypothetical information system. Then 12 of the analysts are instructed to produce the information system by using current technology. The other 12 analysts are trained in the use of the new software package and then instructed to use it to produce the information system.

This study involves two populations: a population of systems analysts using the current technology and a population of systems analysts using the new software package. In terms of the time required to complete the information system design project, the population means are as follow.

$\mu_1$  = the mean project completion time for systems analysts using the current technology

$\mu_2$  = the mean project completion time for systems analysts using the new software package

The researcher in charge of the new software evaluation project hopes to show that the new software package will provide a shorter mean project completion time. Thus, the researcher is looking for evidence to conclude that  $\mu_2$  is less than  $\mu_1$ ; in this case, the

**TABLE 10.1** COMPLETION TIME DATA AND SUMMARY STATISTICS FOR THE SOFTWARE TESTING STUDY

**WEB file**  
SoftwareTest

	Current Technology	New Software
	300	274
	280	220
	344	308
	385	336
	372	198
	360	300
	288	315
	321	258
	376	318
	290	310
	301	332
	283	263
<b>Summary Statistics</b>		
Sample size	$n_1 = 12$	$n_2 = 12$
Sample mean	$\bar{x}_1 = 325$ hours	$\bar{x}_2 = 286$ hours
Sample standard deviation	$s_1 = 40$	$s_2 = 44$

difference between the two population means,  $\mu_1 - \mu_2$ , will be greater than zero. The research hypothesis  $\mu_1 - \mu_2 > 0$  is stated as the alternative hypothesis. Thus, the hypothesis test becomes

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_a: \mu_1 - \mu_2 > 0$$

We will use  $\alpha = .05$  as the level of significance.

Suppose that the 24 analysts complete the study with the results shown in Table 10.1. Using the test statistic in equation (10.8), we have

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(325 - 286) - 0}{\sqrt{\frac{40^2}{12} + \frac{44^2}{12}}} = 2.27$$

Computing the degrees of freedom using equation (10.7), we have

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{40^2}{12} + \frac{44^2}{12}\right)^2}{\frac{1}{12 - 1} \left(\frac{40^2}{12}\right)^2 + \frac{1}{12 - 1} \left(\frac{44^2}{12}\right)^2} = 21.8$$

Rounding down, we will use a  $t$  distribution with 21 degrees of freedom. This row of the  $t$  distribution table is as follows:

Area in Upper Tail	.20	.10	.05	.025	.01	.005
<b><math>t</math>-Value (21 <math>df</math>)</b>	0.859	1.323	1.721	2.080	2.518	2.831

$t = 2.27$

**FIGURE 10.2** MINITAB OUTPUT FOR THE HYPOTHESIS TEST OF THE CURRENT AND NEW SOFTWARE TECHNOLOGY

```

Two-sample T for Current vs New

      N      Mean      StDev      SE Mean
Current  12      325.0      40.0          12
New      12      286.0      44.0          13

Difference = mu Current - mu New
Estimate for difference: 39.0000
95% lower bound for difference = 9.5
T-Test of difference = 0 (vs >):  T-Value = 2.27  P-Value = 0.017  DF = 21

```

Using the  $t$  distribution table, we can only determine a range for the  $p$ -value. Use of Excel or Minitab shows the exact  $p$ -value = .017.

With an upper tail test, the  $p$ -value is the area in the upper tail to the right of  $t = 2.27$ . From the above results, we see that the  $p$ -value is between .025 and .01. Thus, the  $p$ -value is less than  $\alpha = .05$  and  $H_0$  is rejected. The sample results enable the researcher to conclude that  $\mu_1 - \mu_2 > 0$ , or  $\mu_1 > \mu_2$ . Thus, the research study supports the conclusion that the new software package provides a smaller population mean completion time.

Minitab or Excel can be used to analyze data for testing hypotheses about the difference between two population means. The Minitab output comparing the current and new software technology is shown in Figure 10.2. The last line of the output shows  $t = 2.27$  and  $p$ -value = .017. Note that Minitab used equation (10.7) to compute 21 degrees of freedom for this analysis.

### Practical Advice

Whenever possible, equal sample sizes,  $n_1 = n_2$ , are recommended.

The interval estimation and hypothesis testing procedures presented in this section are robust and can be used with relatively small sample sizes. In most applications, equal or nearly equal sample sizes such that the total sample size  $n_1 + n_2$  is at least 20 can be expected to provide very good results even if the populations are not normal. Larger sample sizes are recommended if the distributions of the populations are highly skewed or contain outliers. Smaller sample sizes should only be used if the analyst is satisfied that the distributions of the populations are at least approximately normal.

### NOTES AND COMMENTS

Another approach used to make inferences about the difference between two population means when  $\sigma_1$  and  $\sigma_2$  are unknown is based on the assumption that the two population standard deviations are equal ( $\sigma_1 = \sigma_2 = \sigma$ ). Under this assumption, the two sample standard deviations are combined to provide the following *pooled sample variance*:

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

The  $t$  test statistic becomes

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and has  $n_1 + n_2 - 2$  degrees of freedom. At this point, the computation of the  $p$ -value and the interpretation of the sample results are identical to the procedures discussed earlier in this section.

A difficulty with this procedure is that the assumption that the two population standard deviations are equal is usually difficult to verify. Unequal population standard deviations are frequently encountered. Using the pooled procedure may not provide satisfactory results, especially if the sample sizes  $n_1$  and  $n_2$  are quite different.

The  $t$  procedure that we presented in this section does not require the assumption of equal population standard deviations and can be applied whether the population standard deviations are equal or not. It is a more general procedure and is recommended for most applications.

## Exercises

### Methods

#### SELF test

9. The following results are for independent random samples taken from two populations.

Sample 1	Sample 2
$n_1 = 20$	$n_2 = 30$
$\bar{x}_1 = 22.5$	$\bar{x}_2 = 20.1$
$s_1 = 2.5$	$s_2 = 4.8$

- What is the point estimate of the difference between the two population means?
- What is the degrees of freedom for the  $t$  distribution?
- At 95% confidence, what is the margin of error?
- What is the 95% confidence interval for the difference between the two population means?

#### SELF test

10. Consider the following hypothesis test.

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

$$H_a: \mu_1 - \mu_2 \neq 0$$

The following results are from independent samples taken from two populations.

Sample 1	Sample 2
$n_1 = 35$	$n_2 = 40$
$\bar{x}_1 = 13.6$	$\bar{x}_2 = 10.1$
$s_1 = 5.2$	$s_2 = 8.5$

- What is the value of the test statistic?
  - What is the degrees of freedom for the  $t$  distribution?
  - What is the  $p$ -value?
  - At  $\alpha = .05$ , what is your conclusion?
11. Consider the following data for two independent random samples taken from two normal populations.

<b>Sample 1</b>	10	7	13	7	9	8
<b>Sample 2</b>	8	7	8	4	6	9

- Compute the two sample means.
- Compute the two sample standard deviations.
- What is the point estimate of the difference between the two population means?
- What is the 90% confidence interval estimate of the difference between the two population means?

### Applications

#### SELF test

12. The U.S. Department of Transportation provides the number of miles that residents of the 75 largest metropolitan areas travel per day in a car. Suppose that for a simple random sample of 50 Buffalo residents the mean is 22.5 miles a day and the standard deviation is

8.4 miles a day, and for an independent simple random sample of 40 Boston residents the mean is 18.6 miles a day and the standard deviation is 7.4 miles a day.

- a. What is the point estimate of the difference between the mean number of miles that Buffalo residents travel per day and the mean number of miles that Boston residents travel per day?
- b. What is the 95% confidence interval for the difference between the two population means?



13. FedEx and United Parcel Service (UPS) are the world's two leading cargo carriers by volume and revenue (*The Wall Street Journal*, January 27, 2004). According to the Airports Council International, the Memphis International Airport (FedEx) and the Louisville International Airport (UPS) are 2 of the 10 largest cargo airports in the world. The following random samples show the tons of cargo per day handled by these airports. Data are in thousands of tons.

Memphis					
9.1	15.1	8.8	10.0	7.5	10.5
8.3	9.1	6.0	5.8	12.1	9.3
Louisville					
4.7	5.0	4.2	3.3	5.5	
2.2	4.1	2.6	3.4	7.0	

- a. Compute the sample mean and sample standard deviation for each airport.
  - b. What is the point estimate of the difference between the two population means? Interpret this value in terms of the higher-volume airport and a comparison of the volume difference between the two airports.
  - c. Develop a 95% confidence interval of the difference between the daily population means for the two airports.
14. Are nursing salaries in Tampa, Florida, lower than those in Dallas, Texas? Salary data show staff nurses in Tampa earn less than staff nurses in Dallas (*The Tampa Tribune*, January 15, 2007). Suppose that in a follow-up study of 40 staff nurses in Tampa and 50 staff nurses in Dallas you obtain the following results.

Tampa	Dallas
$n_1 = 40$	$n_2 = 50$
$\bar{x}_1 = \$56,100$	$\bar{x}_2 = \$59,400$
$s_1 = \$6000$	$s_2 = \$7000$

- a. Formulate hypothesis so that, if the null hypothesis is rejected, we can conclude that salaries for staff nurses in Tampa are significantly lower than for those in Dallas. Use  $\alpha = .05$ .
  - b. What is the value of the test statistic?
  - c. What is the  $p$ -value?
  - d. What is your conclusion?
15. Injuries to Major League Baseball players have been increasing in recent years. For the period 1992 to 2001, league expansion caused Major League Baseball rosters to increase 15%. However, the number of players being put on the disabled list due to injury increased 32% over the same period (*USA Today*, July 8, 2002). A research question addressed whether Major League Baseball players being put on the disabled list are on the list longer in 2001 than players put on the disabled list a decade earlier.

- a. Using the population mean number of days a player is on the disabled list, formulate null and alternative hypotheses that can be used to test the research question.
- b. Assume that the following data apply:

	2001 Season	1992 Season
Sample size	$n_1 = 45$	$n_2 = 38$
Sample mean	$\bar{x}_1 = 60$ days	$\bar{x}_2 = 51$ days
Sample standard deviation	$s_1 = 18$ days	$s_2 = 15$ days

What is the point estimate of the difference between population mean number of days on the disabled list for 2001 compared to 1992? What is the percentage increase in the number of days on the disabled list?

- c. Use  $\alpha = .01$ . What is your conclusion about the number of days on the disabled list? What is the  $p$ -value?
  - d. Do these data suggest that Major League Baseball should be concerned about the situation?
16. The College Board provided comparisons of Scholastic Aptitude Test (SAT) scores based on the highest level of education attained by the test taker's parents. A research hypothesis was that students whose parents had attained a higher level of education would on average score higher on the SAT. During 2003, the overall mean SAT verbal score was 507 (*The World Almanac*, 2004). SAT verbal scores for independent samples of students follow. The first sample shows the SAT verbal test scores for students whose parents are college graduates with a bachelor's degree. The second sample shows the SAT verbal test scores for students whose parents are high school graduates but do not have a college degree.



#### Student's Parents

College Grads		High School Grads	
485	487	442	492
534	533	580	478
650	526	479	425
554	410	486	485
550	515	528	390
572	578	524	535
497	448		
592	469		

- a. Formulate the hypotheses that can be used to determine whether the sample data support the hypothesis that students show a higher population mean verbal score on the SAT if their parents attained a higher level of education.
  - b. What is the point estimate of the difference between the means for the two populations?
  - c. Compute the  $p$ -value for the hypothesis test.
  - d. At  $\alpha = .05$ , what is your conclusion?
17. Periodically, Merrill Lynch customers are asked to evaluate Merrill Lynch financial consultants and services. Higher ratings on the client satisfaction survey indicate better service, with 7 the maximum service rating. Independent samples of service ratings for two financial consultants are summarized here. Consultant A has 10 years of experience, whereas consultant B has 1 year of experience. Use  $\alpha = .05$  and test to see whether the consultant with more experience has the higher population mean service rating.

## Consultant A

$$\begin{aligned}n_1 &= 16 \\ \bar{x}_1 &= 6.82 \\ s_1 &= .64\end{aligned}$$

## Consultant B

$$\begin{aligned}n_2 &= 10 \\ \bar{x}_2 &= 6.25 \\ s_2 &= .75\end{aligned}$$

## WEB file

SAT

- State the null and alternative hypotheses.
  - Compute the value of the test statistic.
  - What is the  $p$ -value?
  - What is your conclusion?
18. Educational testing companies provide tutoring, classroom learning, and practice tests in an effort to help students perform better on tests such as the Scholastic Aptitude Test (SAT). The test preparation companies claim that their courses will improve SAT score performances by an average of 120 points (*The Wall Street Journal*, January 23, 2003). A researcher is uncertain of this claim and believes that 120 points may be an overstatement in an effort to encourage students to take the test preparation course. In an evaluation study of one test preparation service, the researcher collects SAT score data for 35 students who took the test preparation course and 48 students who did not take the course. The file named SAT contains the scores for this study.
- Formulate the hypotheses that can be used to test the researcher's belief that the improvement in SAT scores may be less than the stated average of 120 points.
  - Using  $\alpha = .05$ , what is your conclusion?
  - What is the point estimate of the improvement in the average SAT scores provided by the test preparation course? Provide a 95% confidence interval estimate of the improvement.
  - What advice would you have for the researcher after seeing the confidence interval?

## 10.3

## Inferences About the Difference Between Two Population Means: Matched Samples

Suppose employees at a manufacturing company can use two different methods to perform a production task. To maximize production output, the company wants to identify the method with the smaller population mean completion time. Let  $\mu_1$  denote the population mean completion time for production method 1 and  $\mu_2$  denote the population mean completion time for production method 2. With no preliminary indication of the preferred production method, we begin by tentatively assuming that the two production methods have the same population mean completion time. Thus, the null hypothesis is  $H_0: \mu_1 - \mu_2 = 0$ . If this hypothesis is rejected, we can conclude that the population mean completion times differ. In this case, the method providing the smaller mean completion time would be recommended. The null and alternative hypotheses are written as follows.

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

$$H_a: \mu_1 - \mu_2 \neq 0$$

In choosing the sampling procedure that will be used to collect production time data and test the hypotheses, we consider two alternative designs. One is based on independent samples and the other is based on **matched samples**.

- Independent sample design:** A simple random sample of workers is selected and each worker in the sample uses method 1. A second independent simple random sample of workers is selected and each worker in this sample uses method 2. The

test of the difference between population means is based on the procedures in Section 10.2.

2. *Matched sample design:* One simple random sample of workers is selected. Each worker first uses one method and then uses the other method. The order of the two methods is assigned randomly to the workers, with some workers performing method 1 first and others performing method 2 first. Each worker provides a pair of data values, one value for method 1 and another value for method 2.

In the matched sample design the two production methods are tested under similar conditions (i.e., with the same workers); hence this design often leads to a smaller sampling error than the independent sample design. The primary reason is that in a matched sample design, variation between workers is eliminated because the same workers are used for both production methods.

Let us demonstrate the analysis of a matched sample design by assuming it is the method used to test the difference between population means for the two production methods. A random sample of six workers is used. The data on completion times for the six workers are given in Table 10.2. Note that each worker provides a pair of data values, one for each production method. Also note that the last column contains the difference in completion times  $d_i$  for each worker in the sample.

The key to the analysis of the matched sample design is to realize that we consider only the column of differences. Therefore, we have six data values (.6, −.2, .5, .3, .0, and .6) that will be used to analyze the difference between population means of the two production methods.

Let  $\mu_d$  = the mean of the *difference* in values for the population of workers. With this notation, the null and alternative hypotheses are rewritten as follows.

$$H_0: \mu_d = 0$$

$$H_a: \mu_d \neq 0$$

If  $H_0$  is rejected, we can conclude that the population mean completion times differ.

The  $d$  notation is a reminder that the matched sample provides *difference* data. The sample mean and sample standard deviation for the six difference values in Table 10.2 follow.

$$\bar{d} = \frac{\sum d_i}{n} = \frac{1.8}{6} = .30$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n - 1}} = \sqrt{\frac{.56}{5}} = .335$$

*Other than the use of the  $d$  notation, the formulas for the sample mean and sample standard deviation are the same ones used previously in the text.*

**TABLE 10.2** TASK COMPLETION TIMES FOR A MATCHED SAMPLE DESIGN

Worker	Completion Time for Method 1 (minutes)	Completion Time for Method 2 (minutes)	Difference in Completion Times ( $d_i$ )
1	6.0	5.4	.6
2	5.0	5.2	−.2
3	7.0	6.5	.5
4	6.2	5.9	.3
5	6.0	6.0	.0
6	6.4	5.8	.6



*It is not necessary to make the assumption that the population has a normal distribution if the sample size is large. Sample size guidelines for using the  $t$  distribution were presented in Chapters 8 and 9.*

With the small sample of  $n = 6$  workers, we need to make the assumption that the population of differences has a normal distribution. This assumption is necessary so that we may use the  $t$  distribution for hypothesis testing and interval estimation procedures. Based on this assumption, the following test statistic has a  $t$  distribution with  $n - 1$  degrees of freedom.

TEST STATISTIC FOR HYPOTHESIS TESTS INVOLVING MATCHED SAMPLES

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \tag{10.9}$$

*Once the difference data are computed, the  $t$  distribution procedure for matched samples is the same as the one-population estimation and hypothesis testing procedures described in Chapters 8 and 9.*

Let us use equation (10.9) to test the hypotheses  $H_0: \mu_d = 0$  and  $H_a: \mu_d \neq 0$ , using  $\alpha = .05$ . Substituting the sample results  $\bar{d} = .30$ ,  $s_d = .335$ , and  $n = 6$  into equation (10.9), we compute the value of the test statistic.

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{.30 - 0}{.335/\sqrt{6}} = 2.20$$

Now let us compute the  $p$ -value for this two-tailed test. Because  $t = 2.20 > 0$ , the test statistic is in the upper tail of the  $t$  distribution. With  $t = 2.20$ , the area in the upper tail to the right of the test statistic can be found by using the  $t$  distribution table with degrees of freedom =  $n - 1 = 6 - 1 = 5$ . Information from the 5 degrees of freedom row of the  $t$  distribution table is as follows:

Area in Upper Tail	.20	.10	.05	.025	.01	.005
$t$ -Value (5 $df$ )	0.920	1.476	2.015	2.571	3.365	4.032

$t = 2.20$  (with an arrow pointing to the value 2.571 in the table)

Thus, we see that the area in the upper tail is between .05 and .025. Because this test is a two-tailed test, we double these values to conclude that the  $p$ -value is between .10 and .05. This  $p$ -value is greater than  $\alpha = .05$ . Thus, the null hypothesis  $H_0: \mu_d = 0$  is not rejected. Using Excel or Minitab and the data in Table 10.2, we find the exact  $p$ -value = .080.

In addition we can obtain an interval estimate of the difference between the two population means by using the single population methodology of Chapter 8. At 95% confidence, the calculation follows.

$$\begin{aligned} &\bar{d} \pm t_{.025} \frac{s_d}{\sqrt{n}} \\ &.3 \pm 2.571 \left( \frac{.335}{\sqrt{6}} \right) \\ &.3 \pm .35 \end{aligned}$$

Thus, the margin of error is .35 and the 95% confidence interval for the difference between the population means of the two production methods is  $-.05$  minutes to  $.65$  minutes.

## NOTES AND COMMENTS

- In the example presented in this section, workers performed the production task with first one method and then the other method. This example illustrates a matched sample design in which each sampled element (worker) provides a pair of data values. It is also possible to use different but “similar” elements to provide the pair of data values. For example, a worker at one location could be matched with a similar worker at another location (similarity based on age, education, gender, experience, etc.). The pairs of workers would provide the difference data that could be used in the matched sample analysis.
- A matched sample procedure for inferences about two population means generally provides better precision than the independent sample approach; therefore it is the recommended design. However, in some applications the matching cannot be achieved, or perhaps the time and cost associated with matching are excessive. In such cases, the independent sample design should be used.

## Exercises

### Methods

#### SELF test

19. Consider the following hypothesis test.

$$H_0: \mu_d \leq 0$$

$$H_a: \mu_d > 0$$

The following data are from matched samples taken from two populations.

Element	Population	
	1	2
1	21	20
2	28	26
3	18	18
4	20	20
5	26	24

- Compute the difference value for each element.
  - Compute  $\bar{d}$ .
  - Compute the standard deviation  $s_d$ .
  - Conduct a hypothesis test using  $\alpha = .05$ . What is your conclusion?
20. The following data are from matched samples taken from two populations.

Element	Population	
	1	2
1	11	8
2	7	8
3	9	6
4	12	7
5	13	10
6	15	15
7	15	14

- Compute the difference value for each element.
- Compute  $\bar{d}$ .
- Compute the standard deviation  $s_d$ .
- What is the point estimate of the difference between the two population means?
- Provide a 95% confidence interval for the difference between the two population means.

## Applications

### SELF test

21. A market research firm used a sample of individuals to rate the purchase potential of a particular product before and after the individuals saw a new television commercial about the product. The purchase potential ratings were based on a 0 to 10 scale, with higher values indicating a higher purchase potential. The null hypothesis stated that the mean rating “after” would be less than or equal to the mean rating “before.” Rejection of this hypothesis would show that the commercial improved the mean purchase potential rating. Use  $\alpha = .05$  and the following data to test the hypothesis and comment on the value of the commercial.

Individual	Purchase Rating		Individual	Purchase Rating	
	After	Before		After	Before
1	6	5	5	3	5
2	6	4	6	9	8
3	7	7	7	7	5
4	4	3	8	6	6

### WEB file

Earnings2005

22. Per-share earnings data comparing the current quarter’s earnings with the previous quarter are in the file entitled Earnings2005 (*The Wall Street Journal*, January 27, 2006). Provide a 95% confidence interval estimate of the difference between the population mean for the current quarter versus the previous quarter. Have earnings increased?
23. Bank of America’s Consumer Spending Survey collected data on annual credit card charges in seven different categories of expenditures: transportation, groceries, dining out, household expenses, home furnishings, apparel, and entertainment (*US Airways Attaché*, December 2003). Using data from a sample of 42 credit card accounts, assume that each account was used to identify the annual credit card charges for groceries (population 1) and the annual credit card charges for dining out (population 2). Using the difference data, the sample mean difference was  $\bar{d} = \$850$ , and the sample standard deviation was  $s_d = \$1123$ .
- Formulate the null and alternative hypotheses to test for no difference between the population mean credit card charges for groceries and the population mean credit card charges for dining out.
  - Use a .05 level of significance. Can you conclude that the population means differ? What is the  $p$ -value?
  - Which category, groceries or dining out, has a higher population mean annual credit card charge? What is the point estimate of the difference between the population means? What is the 95% confidence interval estimate of the difference between the population means?

### WEB file

AirFare

24. Airline travelers often choose which airport to fly from based on flight cost. Cost data (in dollars) for a sample of flights to eight cities from Dayton, Ohio, and Louisville, Kentucky, were collected to help determine which of the two airports was more costly to fly from (*The Cincinnati Enquirer*, February 19, 2006). A researcher argued that it is significantly more costly to fly out of Dayton than Louisville. Use the sample data to see whether they support the researcher’s argument. Use  $\alpha = .05$  as the level of significance.

Destination	Dayton	Louisville
Chicago O'Hare	\$319	\$142
Grand Rapids, Michigan	192	213
Portland, Oregon	503	317
Atlanta	256	387
Seattle	339	317
South Bend, Indiana	379	167
Miami	268	273
Dallas–Ft. Worth	288	274

25. In recent years, a growing array of entertainment options competes for consumer time. By 2004, cable television and radio surpassed broadcast television, recorded music, and the daily newspaper to become the two entertainment media with the greatest usage (*The Wall Street Journal*, January 26, 2004). Researchers used a sample of 15 individuals and collected data on the hours per week spent watching cable television and hours per week spent listening to the radio.

**WEB file**  
TVRadio

Individual	Television	Radio	Individual	Television	Radio
1	22	25	9	21	21
2	8	10	10	23	23
3	25	29	11	14	15
4	22	19	12	14	18
5	12	13	13	14	17
6	26	28	14	16	15
7	22	23	15	24	23
8	19	21			

- a. Use a .05 level of significance and test for a difference between the population mean usage for cable television and radio. What is the  $p$ -value?
- b. What is the sample mean number of hours per week spent watching cable television? What is the sample mean number of hours per week spent listening to radio? Which medium has the greater usage?
26. Scores in the first and fourth (final) rounds for a sample of 20 golfers who competed in PGA tournaments are shown in the following table (*Golfweek*, February 14, 2009, and February 28, 2009). Suppose you would like to determine if the mean score for the first round of a PGA Tour event is significantly different than the mean score for the fourth and final round. Does the pressure of playing in the final round cause scores to go up? Or does the increased player concentration cause scores to come down?

**WEB file**  
GolfScores

Player	First Round	Final Round	Player	First Round	Final Round
Michael Letzig	70	72	Aron Price	72	72
Scott Verplank	71	72	Charles Howell	72	70
D. A. Points	70	75	Jason Dufner	70	73
Jerry Kelly	72	71	Mike Weir	70	77
Soren Hansen	70	69	Carl Pettersson	68	70
D. J. Trahan	67	67	Bo Van Pelt	68	65
Bubba Watson	71	67	Ernie Els	71	70
Reteif Goosen	68	75	Cameron Beckman	70	68
Jeff Klauk	67	73	Nick Watney	69	68
Kenny Perry	70	69	Tommy Armour III	67	71

- a. Use  $\alpha = .10$  to test for a statistically significant difference between the population means for first- and fourth-round scores. What is the  $p$ -value? What is your conclusion?
  - b. What is the point estimate of the difference between the two population means? For which round is the population mean score lower?
  - c. What is the margin of error for a 90% confidence interval estimate for the difference between the population means? Could this confidence interval have been used to test the hypothesis in part (a)? Explain.
27. A manufacturer produces both a deluxe and a standard model of an automatic sander designed for home use. Selling prices obtained from a sample of retail outlets follow.

Retail Outlet	Model Price (\$)		Retail Outlet	Model Price (\$)	
	Deluxe	Standard		Deluxe	Standard
1	39	27	5	40	30
2	39	28	6	39	34
3	45	35	7	35	29
4	38	30			

- a. The manufacturer's suggested retail prices for the two models show a \$10 price differential. Use a .05 level of significance and test that the mean difference between the prices of the two models is \$10.
- b. What is the 95% confidence interval for the difference between the mean prices of the two models?

## 10.4

## Inferences About the Difference Between Two Population Proportions

Letting  $p_1$  denote the proportion for population 1 and  $p_2$  denote the proportion for population 2, we next consider inferences about the difference between the two population proportions:  $p_1 - p_2$ . To make an inference about this difference, we will select two independent random samples consisting of  $n_1$  units from population 1 and  $n_2$  units from population 2.

### Interval Estimation of $p_1 - p_2$

In the following example, we show how to compute a margin of error and develop an interval estimate of the difference between two population proportions.

A tax preparation firm is interested in comparing the quality of work at two of its regional offices. By randomly selecting samples of tax returns prepared at each office and verifying the sample returns' accuracy, the firm will be able to estimate the proportion of erroneous returns prepared at each office. Of particular interest is the difference between these proportions.

$p_1$  = proportion of erroneous returns for population 1 (office 1)

$p_2$  = proportion of erroneous returns for population 2 (office 2)

$\bar{p}_1$  = sample proportion for a simple random sample from population 1

$\bar{p}_2$  = sample proportion for a simple random sample from population 2

The difference between the two population proportions is given by  $p_1 - p_2$ . The point estimator of  $p_1 - p_2$  is as follows.

POINT ESTIMATOR OF THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

$$\bar{p}_1 - \bar{p}_2 \quad (10.10)$$

Thus, the point estimator of the difference between two population proportions is the difference between the sample proportions of two independent simple random samples.

As with other point estimators, the point estimator  $\bar{p}_1 - \bar{p}_2$  has a sampling distribution that reflects the possible values of  $\bar{p}_1 - \bar{p}_2$  if we repeatedly took two independent random samples. The mean of this sampling distribution is  $p_1 - p_2$  and the standard error of  $\bar{p}_1 - \bar{p}_2$  is as follows:

STANDARD ERROR OF  $\bar{p}_1 - \bar{p}_2$

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (10.11)$$

If the sample sizes are large enough that  $n_1p_1$ ,  $n_1(1-p_1)$ ,  $n_2p_2$ , and  $n_2(1-p_2)$  are all greater than or equal to 5, the sampling distribution of  $\bar{p}_1 - \bar{p}_2$  can be approximated by a normal distribution.

As we showed previously, an interval estimate is given by a point estimate  $\pm$  a margin of error. In the estimation of the difference between two population proportions, an interval estimate will take the following form:

$$\bar{p}_1 - \bar{p}_2 \pm \text{Margin of error}$$

With the sampling distribution of  $\bar{p}_1 - \bar{p}_2$  approximated by a normal distribution, we would like to use  $z_{\alpha/2}\sigma_{\bar{p}_1 - \bar{p}_2}$  as the margin of error. However,  $\sigma_{\bar{p}_1 - \bar{p}_2}$  given by equation (10.11) cannot be used directly because the two population proportions,  $p_1$  and  $p_2$ , are unknown. Using the sample proportion  $\bar{p}_1$  to estimate  $p_1$  and the sample proportion  $\bar{p}_2$  to estimate  $p_2$ , the margin of error is as follows.

$$\text{Margin of error} = z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \quad (10.12)$$

The general form of an interval estimate of the difference between two population proportions is as follows.

INTERVAL ESTIMATE OF THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \quad (10.13)$$

where  $1 - \alpha$  is the confidence coefficient.

Returning to the tax preparation example, we find that independent simple random samples from the two offices provide the following information.

Office 1	Office 2
$n_1 = 250$	$n_2 = 300$
Number of returns with errors = 35	Number of returns with errors = 27



The sample proportions for the two offices follow.

$$\bar{p}_1 = \frac{35}{250} = .14$$

$$\bar{p}_2 = \frac{27}{300} = .09$$

The point estimate of the difference between the proportions of erroneous tax returns for the two populations is  $\bar{p}_1 - \bar{p}_2 = .14 - .09 = .05$ . Thus, we estimate that office 1 has a .05, or 5%, greater error rate than office 2.

Expression (10.13) can now be used to provide a margin of error and interval estimate of the difference between the two population proportions. Using a 90% confidence interval with  $z_{\alpha/2} = z_{.05} = 1.645$ , we have

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2(1 - \bar{p}_2)}{n_2}}$$

$$.14 - .09 \pm 1.645 \sqrt{\frac{.14(1 - .14)}{250} + \frac{.09(1 - .09)}{300}}$$

$$.05 \pm .045$$

Thus, the margin of error is .045, and the 90% confidence interval is .005 to .095.

## Hypothesis Tests About $p_1 - p_2$

Let us now consider hypothesis tests about the difference between the proportions of two populations. We focus on tests involving no difference between the two population proportions. In this case, the three forms for a hypothesis test are as follows:

*All hypotheses considered use 0 as the difference of interest.*

$$\begin{array}{lll} H_0: p_1 - p_2 \geq 0 & H_0: p_1 - p_2 \leq 0 & H_0: p_1 - p_2 = 0 \\ H_a: p_1 - p_2 < 0 & H_a: p_1 - p_2 > 0 & H_a: p_1 - p_2 \neq 0 \end{array}$$

When we assume  $H_0$  is true as an equality, we have  $p_1 - p_2 = 0$ , which is the same as saying that the population proportions are equal,  $p_1 = p_2$ .

We will base the test statistic on the sampling distribution of the point estimator  $\bar{p}_1 - \bar{p}_2$ . In equation (10.11), we showed that the standard error of  $\bar{p}_1 - \bar{p}_2$  is given by

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Under the assumption  $H_0$  is true as an equality, the population proportions are equal and  $p_1 = p_2 = p$ . In this case,  $\sigma_{\bar{p}_1 - \bar{p}_2}$  becomes

STANDARD ERROR OF  $\bar{p}_1 - \bar{p}_2$  WHEN  $p_1 = p_2 = p$ 

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad (10.14)$$

With  $p$  unknown, we pool, or combine, the point estimators from the two samples ( $\bar{p}_1$  and  $\bar{p}_2$ ) to obtain a single point estimator of  $p$  as follows:

POOLED ESTIMATOR OF  $p$  WHEN  $p_1 = p_2 = p$ 

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} \quad (10.15)$$

This **pooled estimator of  $p$**  is a weighted average of  $\bar{p}_1$  and  $\bar{p}_2$ .

Substituting  $\bar{p}$  for  $p$  in equation (10.14), we obtain an estimate of the standard error of  $\bar{p}_1 - \bar{p}_2$ . This estimate of the standard error is used in the test statistic. The general form of the test statistic for hypothesis tests about the difference between two population proportions is the point estimator divided by the estimate of  $\sigma_{\bar{p}_1 - \bar{p}_2}$ :

TEST STATISTIC FOR HYPOTHESIS TESTS ABOUT  $p_1 - p_2$ 

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (10.16)$$

This test statistic applies to large sample situations where  $n_1p_1$ ,  $n_1(1-p_1)$ ,  $n_2p_2$ , and  $n_2(1-p_2)$  are all greater than or equal to 5.

Let us return to the tax preparation firm example and assume that the firm wants to use a hypothesis test to determine whether the error proportions differ between the two offices. A two-tailed test is required. The null and alternative hypotheses are as follows:

$$\begin{aligned} H_0: p_1 - p_2 &= 0 \\ H_a: p_1 - p_2 &\neq 0 \end{aligned}$$

If  $H_0$  is rejected, the firm can conclude that the error rates at the two offices differ. We will use  $\alpha = .10$  as the level of significance.

The sample data previously collected showed  $\bar{p}_1 = .14$  for the  $n_1 = 250$  returns sampled at office 1 and  $\bar{p}_2 = .09$  for the  $n_2 = 300$  returns sampled at office 2. We continue by computing the pooled estimate of  $p$ .

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{250(.14) + 300(.09)}{250 + 300} = .1127$$

Using this pooled estimate and the difference between the sample proportions, the value of the test statistic is as follows.

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(.14 - .09)}{\sqrt{.1127(1 - .1127)\left(\frac{1}{250} + \frac{1}{300}\right)}} = 1.85$$

In computing the  $p$ -value for this two-tailed test, we first note that  $z = 1.85$  is in the upper tail of the standard normal distribution. Using  $z = 1.85$  and the standard normal distribution table, we find the area in the upper tail is  $1.0000 - .9678 = .0322$ . Doubling this area for a two-tailed test, we find the  $p$ -value =  $2(.0322) = .0644$ . With the  $p$ -value less than  $\alpha = .10$ ,  $H_0$  is rejected at the .10 level of significance. The firm can conclude that the error rates differ between the two offices. This hypothesis testing conclusion is consistent with the earlier interval estimation results that showed the interval estimate of the difference between the population error rates at the two offices to be .005 to .095, with Office 1 having the higher error rate.

## Exercises

### Methods

#### SELF test

28. Consider the following results for independent samples taken from two populations.

Sample 1	Sample 2
$n_1 = 400$	$n_2 = 300$
$\bar{p}_1 = .48$	$\bar{p}_2 = .36$

- What is the point estimate of the difference between the two population proportions?
- Develop a 90% confidence interval for the difference between the two population proportions.
- Develop a 95% confidence interval for the difference between the two population proportions.

#### SELF test

29. Consider the hypothesis test

$$H_0: p_1 - p_2 \leq 0$$

$$H_a: p_1 - p_2 > 0$$

The following results are for independent samples taken from the two populations.

Sample 1	Sample 2
$n_1 = 200$	$n_2 = 300$
$\bar{p}_1 = .22$	$\bar{p}_2 = .16$

- What is the  $p$ -value?
- With  $\alpha = .05$ , what is your hypothesis testing conclusion?

## Applications

30. A *BusinessWeek*/Harris survey asked senior executives at large corporations their opinions about the economic outlook for the future. One question was, “Do you think that there will be an increase in the number of full-time employees at your company over the next 12 months?” In the current survey, 220 of 400 executives answered yes, while in a previous year survey, 192 of 400 executives had answered yes. Provide a 95% confidence interval estimate for the difference between the proportions at the two points in time. What is your interpretation of the interval estimate?
31. The Professional Golf Association (PGA) measured the putting accuracy of professional golfers playing on the PGA Tour and the best amateur golfers playing in the World Amateur Championship (*Golf Magazine*, January 2007). A sample of 1075 6-foot putts by professional golfers found 688 made putts. A sample of 1200 6-foot putts by amateur golfers found 696 made putts.
- Estimate the proportion of made 6-foot putts by professional golfers. Estimate the proportion of made 6-foot putts by amateur golfers. Which group had a better putting accuracy?
  - What is the point estimate of the difference between the proportions of the two populations? What does this estimate tell you about the percentage of putts made by the two groups of golfers?
  - What is the 95% confidence interval for the difference between the two population proportions? Interpret his confidence interval in terms of the percentage of putts made by the two groups of golfers.
32. An American Automobile Association (AAA) study investigated the question of whether a man or a woman was more likely to stop and ask for directions (AAA, January 2006). The situation referred to in the study stated the following: “If you and your spouse are driving together and become lost, would you stop and ask for directions?” A sample representative of the data used by AAA showed 300 of 811 women said that they would stop and ask for directions, while 255 of 750 men said that they would stop and ask for directions.
- The AAA research hypothesis was that women would be more likely to say that they would stop and ask for directions. Formulate the null and alternative hypotheses for this study.
  - What is the percentage of women who indicated that they would stop and ask for directions?
  - What is the percentage of men who indicated that they would stop and ask for directions?
  - At  $\alpha = .05$ , test the hypothesis. What is the  $p$ -value, and what conclusion would you expect AAA to draw from this study?
33. Chicago O’Hare and Atlanta Hartsfield-Jackson are the two busiest airports in the United States. The congestion often leads to delayed flight arrivals as well as delayed flight departures. The Bureau of Transportation tracks the on-time and delayed performance at major airports (*Travel & Leisure*, November 2006). A flight is considered delayed if it is more than 15 minutes behind schedule. The following sample data show the delayed departures at Chicago O’Hare and Atlanta Hartsfield-Jackson airports.

	Chicago O’Hare	Atlanta Hartsfield-Jackson
Flights	900	1200
Delayed Departures	252	312

- State the hypotheses that can be used to determine whether the population proportions of delayed departures differ at these two airports.
- What is the point estimate of the proportion of flights that have delayed departures at Chicago O’Hare?

- c. What is the point estimate of the proportion of flights that have delayed departures at Atlanta Hartsfield-Jackson?
- d. What is the  $p$ -value for the hypothesis test? What is your conclusion?
34. *BusinessWeek* reported that there seems to be a difference by age group in how well people like life in Russia (*BusinessWeek*, March 10, 2008). The following sample data are consistent with the *BusinessWeek* findings and show the responses by age group to the question: “Do you like life in Russia?”

	Russian Age Group	
	17–26	40 and over
Sample	300	260
Responded Yes	192	117

- a. What is the point estimate of the proportion of Russians aged 17 to 26 who like life in Russia?
- b. What is the point estimate of the proportion of Russians aged 40 and over who like life in Russia?
- c. Provide a 95% confidence interval estimate of the difference between the proportion of young Russians aged 17 to 26 and older Russians aged 40 and over who like life in Russia.
35. In a test of the quality of two television commercials, each commercial was shown in a separate test area six times over a one-week period. The following week a telephone survey was conducted to identify individuals who had seen the commercials. Those individuals were asked to state the primary message in the commercials. The following results were recorded.

	Commercial A	Commercial B
Number Who Saw Commercial	150	200
Number Who Recalled Message	63	60

- a. Use  $\alpha = .05$  and test the hypothesis that there is no difference in the recall proportions for the two commercials.
- b. Compute a 95% confidence interval for the difference between the recall proportions for the two populations.
36. During the 2003 Super Bowl, Miller Lite Beer’s commercial referred to as “The Miller Lite Girls” ranked among the top three most effective advertisements aired during the Super Bowl (*USA Today*, December 29, 2003). The survey of advertising effectiveness, conducted by *USA Today*’s Ad Track poll, reported separate samples by respondent age group to learn about how the Super Bowl advertisement appealed to different age groups. The following sample data apply to the “The Miller Lite Girls” commercial.

Age Group	Sample Size	Liked the Ad a Lot
Under 30	100	49
30 to 49	150	54

- a. Formulate a hypothesis test that can be used to determine whether the population proportions for the two age groups differ.

- b. What is the point estimate of the difference between the two population proportions?
  - c. Conduct the hypothesis test and report the  $p$ -value. At  $\alpha = .05$ , what is your conclusion?
  - d. Discuss the appeal of the advertisements to the younger and the older age groups. Would the Miller Lite organization find the results of the *USA Today* Ad Track poll encouraging? Explain.
37. A 2003 *New York Times*/CBS News poll sampled 523 adults who were planning a vacation during the next six months and found that 141 were expecting to travel by airplane (*New York Times* News Service, March 2, 2003). A similar survey question in a May 1993 *New York Times*/CBS News poll found that of 477 adults who were planning a vacation in the next six months, 81 were expecting to travel by airplane.
- a. State the hypotheses that can be used to determine whether a significant change occurred in the population proportion planning to travel by airplane over the 10-year period.
  - b. What is the sample proportion expecting to travel by airplane in 2003? In 1993?
  - c. Use  $\alpha = .01$  and test for a significant difference. What is your conclusion?
  - d. Discuss reasons that might provide an explanation for this conclusion.

## Summary

In this chapter we discussed procedures for developing interval estimates and conducting hypothesis tests involving two populations. First, we showed how to make inferences about the difference between two population means when independent simple random samples are selected. We first considered the case where the population standard deviations,  $\sigma_1$  and  $\sigma_2$ , could be assumed known. The standard normal distribution  $z$  was used to develop the interval estimate and served as the test statistic for hypothesis tests. We then considered the case where the population standard deviations were unknown and estimated by the sample standard deviations  $s_1$  and  $s_2$ . In this case, the  $t$  distribution was used to develop the interval estimate and served as the test statistic for hypothesis tests.

Inferences about the difference between two population means were then discussed for the matched sample design. In the matched sample design each element provides a pair of data values, one from each population. The difference between the paired data values is then used in the statistical analysis. The matched sample design is generally preferred to the independent sample design because the matched-sample procedure often improves the precision of the estimate.

Finally, interval estimation and hypothesis testing about the difference between two population proportions were discussed. Statistical procedures for analyzing the difference between two population proportions are similar to the procedures for analyzing the difference between two population means.

## Glossary

**Independent simple random samples** Samples selected from two populations in such a way that the elements making up one sample are chosen independently of the elements making up the other sample.

**Matched samples** Samples in which each data value of one sample is matched with a corresponding data value of the other sample.

**Pooled estimator of  $p$**  An estimator of a population proportion obtained by computing a weighted average of the point estimators obtained from two independent samples.

## Key Formulas

### Point Estimator of the Difference Between Two Population Means

$$\bar{x}_1 - \bar{x}_2 \quad (10.1)$$

### Standard Error of $\bar{x}_1 - \bar{x}_2$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.2)$$

### Interval Estimate of the Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Known

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10.4)$$

### Test Statistic for Hypothesis Tests About $\mu_1 - \mu_2$ : $\sigma_1$ and $\sigma_2$ Known

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (10.5)$$

### Interval Estimate of the Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Unknown

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (10.6)$$

### Degrees of Freedom: $t$ Distribution with Two Independent Random Samples

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} \quad (10.7)$$

### Test Statistic for Hypothesis Tests About $\mu_1 - \mu_2$ : $\sigma_1$ and $\sigma_2$ Unknown

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (10.8)$$

### Test Statistic for Hypothesis Tests Involving Matched Samples

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \quad (10.9)$$

### Point Estimator of the Difference Between Two Population Proportions

$$\bar{p}_1 - \bar{p}_2 \quad (10.10)$$

**Standard Error of  $\bar{p}_1 - \bar{p}_2$** 

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (10.11)$$

**Interval Estimate of the Difference Between Two Population Proportions**

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \quad (10.13)$$

**Standard Error of  $\bar{p}_1 - \bar{p}_2$  when  $p_1 = p_2 = p$** 

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (10.14)$$

**Pooled Estimator of  $p$  when  $p_1 = p_2 = p$** 

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} \quad (10.15)$$

**Test Statistic for Hypothesis Tests About  $p_1 - p_2$** 

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (10.16)$$

**Supplementary Exercises**

38. Safegate Foods, Inc., is redesigning the checkout lanes in its supermarkets throughout the country and is considering two designs. Tests on customer checkout times conducted at two stores where the two new systems have been installed result in the following summary of the data.

System A	System B
$n_1 = 120$	$n_2 = 100$
$\bar{x}_1 = 4.1$ minutes	$\bar{x}_2 = 3.4$ minutes
$\sigma_1 = 2.2$ minutes	$\sigma_2 = 1.5$ minutes

Test at the .05 level of significance to determine whether the population mean checkout times of the two systems differ. Which system is preferred?

39. Home values tend to increase over time under normal conditions, but the recession of 2008 and 2009 has reportedly caused the sales price of existing homes to fall nationwide (*BusinessWeek*, March 9, 2009). You would like to see if the data support this conclusion. The file HomePrices contains data on 30 existing home sales in 2006 and 40 existing home sales in 2009.

- a. Provide a point estimate of the difference between the population mean prices for the two years.
  - b. Develop a 99% confidence interval estimate of the difference between the resale prices of houses in 2006 and 2009.
  - c. Would you feel justified in concluding that resale prices of existing homes have declined from 2006 to 2009? Why or why not?
40. Mutual funds are classified as *load* or *no-load* funds. Load funds require an investor to pay an initial fee based on a percentage of the amount invested in the fund. The no-load funds do not require this initial fee. Some financial advisors argue that the load mutual funds may be worth the extra fee because these funds provide a higher mean rate of return than the no-load mutual funds. A sample of 30 load mutual funds and a sample of 30 no-load mutual funds were selected. Data were collected on the annual return for the funds over a five-year period. The data are contained in the data set Mutual. The data for the first five load and first five no-load mutual funds are as follows.



Mutual Funds—Load	Return	Mutual Funds—No Load	Return
American National Growth	15.51	Amana Income Fund	13.24
Arch Small Cap Equity	14.57	Berger One Hundred	12.13
Bartlett Cap Basic	17.73	Columbia International Stock	12.17
Calvert World International	10.31	Dodge & Cox Balanced	16.06
Colonial Fund A	16.23	Evergreen Fund	17.61

- a. Formulate  $H_0$  and  $H_a$  such that rejection of  $H_0$  leads to the conclusion that the load mutual funds have a higher mean annual return over the five-year period.
  - b. Use the 60 mutual funds in the data set Mutual to conduct the hypothesis test. What is the  $p$ -value? At  $\alpha = .05$ , what is your conclusion?
41. The National Association of Home Builders provided data on the cost of the most popular home remodeling projects. Sample data on cost in thousands of dollars for two types of remodeling projects are as follows.

Kitchen	Master Bedroom	Kitchen	Master Bedroom
25.2	18.0	23.0	17.8
17.4	22.9	19.7	24.6
22.8	26.4	16.9	21.0
21.9	24.8	21.8	
19.7	26.9	23.6	

- a. Develop a point estimate of the difference between the population mean remodeling costs for the two types of projects.
  - b. Develop a 90% confidence interval for the difference between the two population means.
42. In early 2009, the economy was experiencing a recession. But how was the recession affecting the stock market? Shown are data from a sample of 15 companies. Shown for each company is the price per share of stock on January 1 and April 30 (*The Wall Street Journal*, May 1, 2009).



Company	January 1 (\$)	April 30 (\$)
Applied Materials	10.13	12.21
Bank of New York	28.33	25.48
Chevron	73.97	66.10
Cisco Systems	16.30	19.32
Coca-Cola	45.27	43.05
Comcast	16.88	15.46
Ford Motors	2.29	5.98
General Electric	16.20	12.65
Johnson & Johnson	59.83	52.36
JP Morgan Chase	31.53	33.00
Microsoft	19.44	20.26
Oracle	17.73	19.34
Pfizer	17.71	13.36
Philip Morris	43.51	36.18
Procter & Gamble	61.82	49.44

- What is the change in the mean price per share of stock over the four-month period?
  - Provide a 90% confident interval estimate of the change in the mean price per share of stock. Interpret the results.
  - What was the percentage change in the mean price per share of stock over the four-month period?
  - If this same percentage change were to occur for the next four months and again for the four months after that, what would be the mean price per share of stock at the end of the year 2009?
43. Jupiter Media used a survey to determine how people use their free time. Watching television was the most popular activity selected by both men and women (*The Wall Street Journal*, January 26, 2004). The proportion of men and the proportion of women who selected watching television as their most popular leisure time activity can be estimated from the following sample data.

Gender	Sample Size	Watching Television
Men	800	248
Women	600	156

- State the hypotheses that can be used to test for a difference between the proportion for the population of men and the proportion for the population of women who selected watching television as their most popular leisure time activity.
  - What is the sample proportion of men who selected watching television as their most popular leisure time activity? What is the sample proportion of women?
  - Conduct the hypothesis test and compute the  $p$ -value. At a .05 level of significance, what is your conclusion?
  - What is the margin of error and 95% confidence interval estimate of the difference between the population proportions?
44. A large automobile insurance company selected samples of single and married male policyholders and recorded the number who made an insurance claim over the preceding three-year period.

Single Policyholders	Married Policyholders
$n_1 = 400$	$n_2 = 900$
Number making claims = 76	Number making claims = 90

- a. Use  $\alpha = .05$ . Test to determine whether the claim rates differ between single and married male policyholders.
  - b. Provide a 95% confidence interval for the difference between the proportions for the two populations.
45. Medical tests were conducted to learn about drug-resistant tuberculosis. Of 142 cases tested in New Jersey, 9 were found to be drug-resistant. Of 268 cases tested in Texas, 5 were found to be drug-resistant. Do these data suggest a statistically significant difference between the proportions of drug-resistant cases in the two states? Use a .02 level of significance. What is the  $p$ -value, and what is your conclusion?
46. Vacation occupancy rates were expected to be up during March 2008 in Myrtle Beach, South Carolina (*The Sun News*, February 29, 2008). Data in the file Occupancy will allow you to replicate the findings presented in the newspaper. The data show units rented and not rented for a random sample of vacation properties during the first week of March 2007 and March 2008.
- a. Estimate the proportion of units rented during the first week of March 2007 and the first week of March 2008.
  - b. Provide a 95% confidence interval for the difference in proportions.
  - c. On the basis of your findings, does it appear March rental rates for 2008 will be up from those a year earlier?
47. For the week ended January 15, 2009, the bullish sentiment of individual investors was 27.6% (*AII Journal*, February 2009). The bullish sentiment was reported to be 48.7% one week earlier and 39.7% one month earlier. The sentiment measures are based on a poll conducted by the American Association of Individual Investors. Assume that each of the bullish sentiment measures was based on a sample size of 240.
- a. Develop a 95% confidence interval for the difference between the bullish sentiment measures for the most recent two weeks.
  - b. Develop hypotheses so that rejection of the null hypothesis will allow us to conclude that the most recent bullish sentiment is weaker than that of one month ago.
  - c. Conduct a hypotheses test of part (b) using  $\alpha = .01$ . What is your conclusion?



## Case Problem Par, Inc.

Par, Inc., is a major manufacturer of golf equipment. Management believes that Par's market share could be increased with the introduction of a cut-resistant, longer-lasting golf ball. Therefore, the research group at Par has been investigating a new golf ball coating designed to resist cuts and provide a more durable ball. The tests with the coating have been promising.

One of the researchers voiced concern about the effect of the new coating on driving distances. Par would like the new cut-resistant ball to offer driving distances comparable to those of the current-model golf ball. To compare the driving distances for the two balls, 40 balls of both the new and current models were subjected to distance tests. The testing was performed with a mechanical hitting machine so that any difference between the mean distances for the two models could be attributed to a difference in the two models. The results of the tests, with distances measured to the nearest yard, follow. These data are available on the website that accompanies the text.

**WEB file**  
Golf

Model		Model		Model		Model	
Current	New	Current	New	Current	New	Current	New
264	277	270	272	263	274	281	283
261	269	287	259	264	266	274	250
267	263	289	264	284	262	273	253
272	266	280	280	263	271	263	260
258	262	272	274	260	260	275	270
283	251	275	281	283	281	267	263
258	262	265	276	255	250	279	261
266	289	260	269	272	263	274	255
259	286	278	268	266	278	276	263
270	264	275	262	268	264	262	279

### Managerial Report

1. Formulate and present the rationale for a hypothesis test that Par could use to compare the driving distances of the current and new golf balls.
2. Analyze the data to provide the hypothesis testing conclusion. What is the  $p$ -value for your test? What is your recommendation for Par, Inc.?
3. Provide descriptive statistical summaries of the data for each model.
4. What is the 95% confidence interval for the population mean driving distance of each model, and what is the 95% confidence interval for the difference between the means of the two populations?
5. Do you see a need for larger sample sizes and more testing with the golf balls? Discuss.

## Appendix 10.1 Inferences About Two Populations Using Minitab

We describe the use of Minitab to develop interval estimates and conduct hypothesis tests about the difference between two population means and the difference between two population proportions. Minitab provides both interval estimation and hypothesis testing results within the same module. Thus, the Minitab procedure is the same for both types of inferences. In the examples that follow, we will demonstrate interval estimation and hypothesis testing for the same two samples. We note that Minitab does not provide a routine for inferences about the difference between two population means when the population standard deviations  $\sigma_1$  and  $\sigma_2$  are known.

### Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Unknown

We will use the data for the checking account balances example presented in Section 10.2. The checking account balances at the Cherry Grove branch are in column C1, and the checking account balances at the Beechmont branch are in column C2. In this example, we will use the Minitab 2-Sample  $t$  procedure to provide a 95% confidence interval estimate of the difference between population means for the checking account balances at the two branch banks. The output of the procedure also provides the  $p$ -value for the hypothesis test:  $H_0: \mu_1 - \mu_2 = 0$  versus  $H_a: \mu_1 - \mu_2 \neq 0$ . The following steps are necessary to execute the procedure:

**WEB file**  
CheckAcct

- Step 1.** Select the **Stat** menu
- Step 2.** Choose **Basic Statistics**
- Step 3.** Choose **2-Sample t**

**Step 4.** When the 2-Sample t (Test and Confidence Interval) dialog box appears:

Select **Samples in different columns**

Enter C1 in the **First** box

Enter C2 in the **Second** box

Select **Options**

**Step 5.** When the 2-Sample t - Options dialog box appears:

Enter 95 in the **Confidence level** box

Enter 0 in the **Test difference** box

Enter not equal in the **Alternative** box

Click **OK**

**Step 6.** Click **OK**

The 95% confidence interval estimate is \$37 to \$193, as described in Section 10.2. The  $p$ -value = .005 shows the null hypothesis of equal population means can be rejected at the  $\alpha = .01$  level of significance. In other applications, step 5 may be modified to provide different confidence levels, different hypothesized values, and different forms of the hypotheses.

## Difference Between Two Population Means with Matched Samples



We use the data on production times in Table 10.2 to illustrate the matched-sample procedure. The completion times for method 1 are entered into column C1 and the completion times for method 2 are entered into column C2. The Minitab steps for a matched sample are as follows:

**Step 1.** Select the **Stat** menu

**Step 2.** Choose **Basic Statistics**

**Step 3.** Choose **Paired t**

**Step 4.** When the Paired t (Test and Confidence Interval) dialog box appears:

Select **Samples in columns**

Enter C1 in the **First sample** box

Enter C2 in the **Second sample** box

Select **Options**

**Step 5.** When the Paired t - Options dialog box appears:

Enter 95 in the **Confidence level**

Enter 0 in the **Test mean** box

Enter not equal in the **Alternative** box

Click **OK**

**Step 6.** Click **OK**

The 95% confidence interval estimate is  $-.05$  to  $.65$ , as described in Section 10.3. The  $p$ -value = .08 shows that the null hypothesis of no difference in completion times cannot be rejected at  $\alpha = .05$ . Step 5 may be modified to provide different confidence levels, different hypothesized values, and different forms of the hypothesis.

## Difference Between Two Population Proportions



We will use the data on tax preparation errors presented in Section 10.4. The sample results for 250 tax returns prepared at Office 1 are in column  $C_1 - T$  and the sample results for 300 tax returns prepared at Office 2 are in column  $C_2 - T$ . Yes denotes an error was found in the tax return and No indicates no error was found. The procedure we describe provides both a 95% confidence interval estimate of the difference between the two population proportions and hypothesis testing results for  $H_0: p_1 - p_2 = 0$  and  $H_a: p_1 - p_2 \neq 0$ .

**Step 1.** Select the **Stat** menu

**Step 2.** Choose **Basic Statistics**

**Step 3. Choose 2 Proportions**

**Step 4.** When the 2 Proportions (Test and Confidence Interval) dialog box appears:

Select **Samples in different columns**

Enter C1 in the **First** box

Enter C2 in the **Second** box

Select **Options**

**Step 5.** When the 2 Proportions-Options dialog box appears:

Enter 90 in the **Confidence level** box

Enter 0 in the **Test difference** box

Enter not equal in the **Alternative** box

Select **Use pooled estimate of p for test**

Click **OK**

**Step 6.** Click **OK**

The 90% confidence interval estimate is .005 to .095, as described in Section 10.4. The  $p$ -value = .065 shows the null hypothesis of no difference in error rates can be rejected at  $\alpha = .10$ . Step 5 may be modified to provide different confidence levels, different hypothesized values, and different forms of the hypotheses.

In the tax preparation example, the data are categorical. Yes and No are used to indicate whether an error is present. In modules involving proportions, Minitab calculates proportions for the response coming second in alphabetic order. Thus, in the tax preparation example, Minitab computes the proportion of Yes responses, which is the proportion we wanted.

If Minitab's alphabetical ordering does not compute the proportion for the response of interest, we can fix it. Select any cell in the data column, go to the Minitab menu bar, and select Editor > Column > Value Order. This sequence will provide the option of entering a user-specified order. Simply make sure that the response of interest is listed second in the define-an-order box. Minitab's 2 Proportion routine will then provide the confidence interval and hypothesis testing results for the population proportion of interest.

Finally, we note that Minitab's 2 Proportion routine uses a computational procedure different from the procedure described in the text. Thus, the Minitab output can be expected to provide slightly different interval estimates and slightly different  $p$ -values. However, results from the two methods should be close and are expected to provide the same interpretation and conclusion.

## Appendix 10.2 Inferences About Two Populations Using Excel

We describe the use of Excel to conduct hypothesis tests about the difference between two population means.\* We begin with inferences about the difference between the means of two populations when the population standard deviations  $\sigma_1$  and  $\sigma_2$  are known.

### Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Known

We will use the examination scores for the two training centers discussed in Section 10.1. The label Center A is in cell A1 and the label Center B is in cell B1. The examination scores for Center A are in cells A2:A31 and examination scores for Center B are in cells B2:B41. The population standard deviations are assumed known with  $\sigma_1 = 10$  and  $\sigma_2 = 10$ . The Excel routine will request the input of variances which are  $\sigma_1^2 = 100$  and  $\sigma_2^2 = 100$ .



\*Excel's data analysis tools provide hypothesis testing procedures for the difference between two population means. No routines are available for interval estimation of the difference between two population means nor for inferences about the difference between two population proportions.

The following steps can be used to conduct a hypothesis test about the difference between the two population means.

- Step 1.** Click the **Data** tab on the Ribbon
- Step 2.** In the **Analysis** group, click **Data Analysis**
- Step 3.** When the Data Analysis dialog box appears:
  - Choose **z-Test: Two Sample for Means**
  - Click **OK**
- Step 4.** When the z-Test: Two Sample for Means dialog box appears:
  - Enter A1:A31 in the **Variable 1 Range** box
  - Enter B1:B41 in the **Variable 2 Range** box
  - Enter 0 in the **Hypothesized Mean Difference** box
  - Enter 100 in the **Variable 1 Variance (known)** box
  - Enter 100 in the **Variable 2 Variance (known)** box
  - Select **Labels**
  - Enter .05 in the **Alpha** box
  - Select **Output Range** and enter C1 in the box
  - Click **OK**

The two-tailed  $p$ -value is denoted  $P(Z \leq z)$  two-tail. Its value of .0977 does not allow us to reject the null hypothesis at  $\alpha = .05$ .

### Difference Between Two Population Means: $\sigma_1$ and $\sigma_2$ Unknown



We use the data for the software testing study in Table 10.1. The data are already entered into an Excel worksheet with the label Current in cell A1 and the label New in cell B1. The completion times for the current technology are in cells A2:A13, and the completion times for the new software are in cells B2:B13. The following steps can be used to conduct a hypothesis test about the difference between two population means with  $\sigma_1$  and  $\sigma_2$  unknown.

- Step 1.** Click the **Data** tab on the Ribbon
- Step 2.** In the **Analysis** group, click **Data Analysis**
- Step 3.** When the Data Analysis dialog box appears:
  - Choose **t-Test: Two Sample Assuming Unequal Variances**
  - Click **OK**
- Step 4.** When the t-Test: Two Sample Assuming Unequal Variances dialog box appears:
  - Enter A1:A13 in the **Variable 1 Range** box
  - Enter B1:B13 in the **Variable 2 Range** box
  - Enter 0 in the **Hypothesized Mean Difference** box
  - Select **Labels**
  - Enter .05 in the **Alpha** box
  - Select **Output Range** and enter C1 in the box
  - Click **OK**

The appropriate  $p$ -value is denoted  $P(T \leq t)$  one-tail. Its value of .017 allows us to reject the null hypothesis at  $\alpha = .05$ .



### Difference Between Two Population Means with Matched Samples

We use the matched-sample completion times in Table 10.2 to illustrate. The data are entered into a worksheet with the label Method 1 in cell A1 and the label Method 2 in cell B2.

The completion times for method 1 are in cells A2:A7 and the completion times for method 2 are in cells B2:B7. The Excel procedure uses the steps previously described for the  $t$ -Test except the user chooses the **t-Test: Paired Two Sample for Means** data analysis tool in step 3. The variable 1 range is A1:A7 and the variable 2 range is B1:B7.

The appropriate  $p$ -value is denoted  $P(T \leq t)$  two-tail. Its value of .08 does not allow us to reject the null hypothesis at  $\alpha = .05$ .

## Appendix 10.3 Inferences About Two Populations Using StatTools

In this appendix we show how StatTools can be used to develop interval estimates and conduct hypothesis tests about the difference between two population means for the  $\sigma_1$  and  $\sigma_2$  unknown case.

### Interval Estimation of $\mu_1$ and $\mu_2$



We will use the data for the checking account balances example presented in Section 10.2. Begin by using the Data Set Manager to create a StatTools data set for these data using the procedure described in the appendix in Chapter 1. The following steps can be used to compute a 95% confidence interval estimate of the difference between the two population means.

- Step 1.** Click the **StatTools** tab on the Ribbon
- Step 2.** In the **Analysis** group, click **Statistical Inference**
- Step 3.** Select the **Confidence Interval** option
- Step 4.** Choose Mean/Std. Deviation
- Step 5.** When the StatTools—Confidence Interval for Mean/Std. Deviation dialog box appears:
  - For **Analysis Type**, choose **Two-Sample Analysis**
  - In the **Variables** section,
    - Select **Cherry Grove**
    - Select **Beechmont**
  - In the **Confidence Intervals to Calculate** section,
    - Select the **For the Difference of Means** option
    - Select 95% for the **Confidence Level**
  - Click **OK**

Because the sample size for Cherry Grove ( $n_1 = 28$ ) differs from the sample size for Beechmont ( $n_2 = 22$ ), StatTools will inform you of this difference after you click OK in step 4. A dialog box will appear saying “The variable Beechmont contains missing data, which this analysis will ignore.” Click OK. A Choose Variable Ordering dialog box then appears, indicating that the analysis will compare the difference between the Cherry Grove data set and the Beechmont data set. Click OK and the StatTools interval estimation output will appear.

### Hypothesis Tests About $\mu_1$ and $\mu_2$



We will use the software evaluation example and the completion time data presented in Table 10.1. Begin by using the Data Set Manager to create a StatTools data set for these data using the procedure described in the appendix in Chapter 1. The following steps can be used to test the hypothesis:  $H_0: \mu_1 - \mu_2 \leq 0$  against  $H_a: \mu_1 - \mu_2 > 0$ .

- Step 1.** Click the **StatTools** tab on the Ribbon
- Step 2.** In the **Analysis** group, click **Statistical Inference**

- Step 3.** Select the **Hypothesis Test** option
- Step 4.** Choose Mean/Std. Deviation
- Step 5.** When the StatTools—Hypothesis Test for Mean/Std. Deviation dialog box appears:
- For **Analysis Type**, choose **Two-Sample Analysis**
  - In the **Variables** section,
    - Select **Current**
    - Select **New**
  - In the **Hypothesis Test to Perform** section,
    - Select **Difference of Means**
    - Enter 0 in the **Null Hypothesis Value** box
    - Select **Greater Than Null Value (One-Tailed Test)** in the **Alternative Hypothesis** box
  - Click **OK**
  - When the Choose Variable Ordering dialog box appears, click **OK**

The results of the hypothesis test will then appear.

## Inferences About the Difference Between Two Population Means: Matched Samples



StatTools can be used to develop interval estimates and conduct hypothesis tests for the difference between population means for the matched samples case. We will use the matched-sample completion times in Table 10.2 to illustrate.

Begin by using the Data Set Manager to create a StatTools data set for these data using the procedure described in the appendix in Chapter 1. The following steps can be used to compute a 95% confidence interval estimate of the difference between the population mean completion times.

- Step 1.** Click the **StatTools** tab on the Ribbon
- Step 2.** In the **Analysis** group, click **Statistical Inference**
- Step 3.** Select the **Confidence Interval** option
- Step 4.** Choose Mean/Std. Deviation
- Step 5.** When the StatTools—Confidence Interval for Mean/Std. Deviation dialog box appears:
- For **Analysis Type**, choose **Paired-Sample Analysis**
  - In the **Variables** section,
    - Select **Method 1**
    - Select **Method 2**
  - In the **Confidence Intervals to Calculate** section,
    - Select the **For the Difference of Means** option
    - Select 95% for the **Confidence Level**
    - If selected, remove the check in the **For the Standard Deviation** box
  - Click **OK**
  - When the Choose Variable Ordering dialog box appears, click **OK**

The confidence interval will appear.

Conducting hypothesis tests for the matched samples case is very similar to conducting hypothesis tests for the difference in two means shown previously. After selecting the Hypothesis Test option in step 3, select the Paired-Sample Analysis option in step 4.